Poisson and Traffic

THE ENO FOUNDATION FOR HIGHWAY TRAFFIC CONTROL
SAUGATUCK · 1955 · CONNECTICUT

Use of Poisson Distribution in Highway Traffic

55-1140.2

Daniel L. Gerlough

The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways

André Schuhl

THE ENO FOUNDATION FOR HIGHWAY TRAFFIC CONTROL
SAUGATUCK · 1955 · CONNECTICUT

Library of Congress catalog card: 55-11902

Printed in the United States of America by Columbia University Press

Foreword

These two mathematical discussions of traffic analyses tend to confirm the theory that if we are properly to understand traffic we must learn more about its behavior; and that if we expect to control traffic efficiently and provide suitable means for control, we must be able to uniformly predict traffic patterns.

To know what to expect will provide a guide to a rule of action to fit the problem.

The authors, Mr. Gerlough and Mr. Schuhl, live in distantly separated parts of the world. It is safe to assume they were motivated in preparing their articles by their daily interest and activities in traffic.

They are highly competent and have that very valuable advantage of experience so essential in translating theory to practice.

Continued efforts in this direction tend to provide an accumulation of basic knowledge of traffic performance and in time to evolve new concepts of our traffic problem with improved ideas in planning and control.

Trusting that this brochure will be some encouragement to further research, the Eno Foundation welcomes the opportunity to publish it and wishes to express its appreciation to the authors.

ENO FOUNDATION

Use of Poisson Distribution in Highway Traffic

Daniel L. Gerlough

Mr. Gerlough, a Registered Professional Electrical Engineer in the state of California, is an engineer with the Institute of Transportation and Traffic Engineering at the University of California at Los Angeles. For ten years prior to his joining the Institute at Los Angeles in 1948, he served as an engineer for six corporations in Texas and California. Mr. Gerlough is a member of the Institute of Radio Engineers, of the American Institute of Electrical Engineers, and of the Institute of Traffic Engineers.

Applications of probability and statistics to engineering problems have increased markedly during the past decade. These techniques are particularly helpful in the study of vehicular traffic. A monograph[1] published recently by the Eno Foundation presents applications of the broad field of statistics to traffic problems. The present paper discusses in detail one particular probability distribution which has been found useful in the treatment of several types of problems encountered by the traffic engineer.

The Poisson distribution is a mathematical relationship which finds such applications as:

Testing the randomness of a given set of data
Fitting of empirical data to a theoretical curve
Prediction of certain phenomena from basic data

[1] Greenshields, Bruce D., and Weida, Frank M., *Statistics with Applications to Highway Traffic Analyses,* The Eno Foundation, 1952.

Some of the uses open to the traffic engineer include:

- Analysis of arrival rates at a given point
- Studies of vehicle spacing (gaps)
- Determination of the probability of finding a vacant parking space
- Studies of certain accidents
- Design of left-turn pockets

The application of the Poisson distribution to traffic problems is not new. Certain applications were discussed by Kinzer[2] in 1933, Adams[3] in 1936, and Greenshields[4] in 1947. It is the purpose here to provide a more extensive treatment in a form that will serve as a reference guide for both students and practitioners of traffic engineering. In attempting to fulfill this objective the following material has been included:

Examples which cover typical cases and at the same time indicate the variety of applicability.

Detailed descriptions of the steps to be followed in applying the Poisson distribution and in testing it by the χ^2 (chi-square) test.

Charts for simplified computations as well as references to standard tables and other computational aids.

Derivation of the Poisson distribution (in sections B to E of the Appendix) starting from the fundamental concepts of permutations and combinations.

Bibliographical references to more advanced applications in which the Poisson distribution is combined with other relationships.

[2] Kinzer, John P., *Application of the Theory of Probability to Problems of Highway Traffic,* thesis submitted in partial satisfaction of requirements for degree of B.C.E., Polytechnic Institute of Brooklyn, June 1, 1933.

Abstracted in: *Proceedings,* Institute of Traffic Engineers, vol. 5, 1934, pp. 118-124.

[3] Adams, William F., "Road Traffic Considered as a Random Series," *Journal* Institution of Civil Engineers, vol. 4, Nov., 1936, pp. 121-130+.

[4] Greenshields, Bruce D.; Shapiro, Donald; Ericksen, Elroy L., *Traffic Performance at Urban Street Intersections,* Technical Report No. 1, Yale Bureau of Highway Traffic, 1947.

Nature of Poisson Distribution

When events of a given group occur in discrete degrees (heads or tails, 1 to 6 on the face of a die, etc.) the probability of occurrence of a particular event in a specified number of trials may be described by the Bernoulli or binomial distribution (See Appendix, Sections B and C).

As an example, let us consider an experiment consisting of five successive drawings of a ball from an urn containing uniformly mixed black and white balls, with the drawn ball being returned to the urn after each drawing. Of the five drawings which comprise a single experiment, let x be the number which produced black balls; thus x can equal 0, 1, 2, 3, 4, or 5. Let $P(x)$ be the probability that in a given experiment the number of black balls drawn would be exactly x. If p is the probability that a particular drawing will yield a black ball, and $q(q = 1-p)$ is the probability that a particular drawing yields a white ball, then

$$P(x) = C_x^5 p^x q^{5-x}$$

where C_x^5 is the number of combinations of 5 things taken x at a time. It can be seen that in the above example p, the probability that a single drawing yields a black ball, is equal to the percentage of black balls in the urn. With the ball being replaced after each drawing p will remain constant from drawing to drawing.

The experiment just described is an example of the "Bernoulli" or "binomial" distribution in which $P(x)$, the probability of exactly x successes out of n trials of an event where the probability of success remains constant from event to event, is given by

$$P(x) = C_x^n p^x q^{n-x}$$

If the number of items in the sample n, becomes very large while the product $pn = m$ is a finite constant, the binomial

distribution approaches the Poisson distribution as a limit.*

$$\lim_{\substack{n \to \infty \\ pn = m}} P(x) = \frac{m^x e^{-m}}{x!}$$

where e is the Napierian base of logarithms.

The derivation of the Poisson distribution as a limiting case of the binomial distribution is given in Appendix D. (The Poisson distribution can be derived independently of the binomial distribution by more advanced concepts. See Fry.[5])

The mathematical conditions of an infinite number of trials and an infinitesimal probability are never achieved in practical problems. Nevertheless, the Poisson distribution is useful as approximating the binomial under appropriate conditions.

For such practical purposes, then, the Poisson distribution may be stated as follows:

If in a given experiment the number of opportunities for an event to occur is large (e.g., $n \geqq 50$)

and

If the number of times a particular event occurs is small (e.g., $P \leqq 0.1$)

and

If the average number of times the event occurs has a finite value, m, ($m = np$),

$$\text{Then} \quad P(x) = \frac{m^x e^{-m}}{x!}$$

where $x = 0, 1, 2, \ldots$

In this statement of the Poisson distribution an experiment may consist of such things as:

a. Observing the number of micro-organisms in a standard sample of blood, x representing the number of micro-organisms in any one sample, and n representing the number of samples studied.
b. Observing the number of alpha particles emitted during each

* This implies that the probability of occurrence, p, becomes very small.

[5] Fry, Thornton C., *Probability and Its Engineering Uses,* Van Nostrand, 1928, pp. 220-227.

successive interval of t seconds. The number of such intervals will be n, and $x_1, x_2 \ldots x_n$ will be the number of particles during the 1st, 2nd, . . . nth intervals.

c. Observing the number of blowholes in each of n castings, x representing the number of holes in any one casting.

Examples of Poisson Distributions

The first record of the use of the Poisson distribution to treat populations having the properties listed is attributed to Bortkewitsch who studied the frequency of death due to the "kick of a horse" among the members of ten Prussian Cavalry Corps during a period of 20 years. Some of the earliest engineering problems treated by the Poisson distribution were telephone problems. The following example is based on such data:[6]

Example 1

CONNECTIONS TO WRONG NUMBER

Number of wrong connections per period	*Number of periods exhibiting the given number of wrong connections (observed frequency)*
0	0
1	0
2	1
3	5
4	11
5	14
6	22
7	43
8	31
9	40
10	35
11	20
12	18
13	12
14	7
15	6
16	2

6 Thorndike, Frances, "Applications of Poisson's Probability Summation," *Bell System Technical Journal,* vol. 5, no. 4, Oct., 1926, pp. 604-624.

The fitting of the Poisson distribution to the experimental data may be carried out in tabular form as follows:

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
$x =$ *Number of wrong connections per period*	*Observed frequency (periods)*	*Total wrong connections*	$P(x) =$ *Probability*	*Theoretical frequency*
0	0	0	0.000	0.0
1	0	0	0.001	0.3
2	1	2	0.006	1.6
3	5	15	0.018	4.8
4	11	44	0.039	10.4
5	14	70	0.068	18.2
6	22	132	0.099	26.4
7	43	301	0.124	33.1
8	31	248	0.135	36.0
9	40	360	0.132	35.2
10	35	350	0.115	30.7
11	20	220	0.091	24.3
12	18	216	0.067	17.9
13	12	156	0.045	12.0
14	7	98	0.028	7.5
15	6	90	0.016	4.3
16	2	32	0.009	2.4
> 16	0	0	0.007	1.9
TOTAL	267	2334	1.000	267.0

The entries in Col. 3 are the products of the corresponding entries in Col. 1 and Col. 2.

m = average number of wrong connections per period =

$$\frac{2334}{267} = 8.742$$

$$P(x) = \frac{(8.742)^x e^{-8.742}}{x!}$$

In the last row of the table (for more than 16 wrong connections per period) the values of probability and theoretical frequency are necessary to make the totals balance. These values represent the summation from 17 to infinity.

Methods of evaluating e^{-x}, $x!$, and $P(x)$ are discussed later in the paper.

$$\text{Calculated Frequency} = (\text{Total periods observed})\ P(x) = 267\ P(x)$$

It will be seen that there is a high degree of agreement between the observed and calculated frequencies.

Note: Column 5 shows the theoretical or calculated frequencies. When these frequencies are calculated, fractional (decimal) values often result. Here these values have been rounded-off to the nearest 0.1. The observed frequencies will, of course, always be integral numbers.

Following the pioneer work in the field of telephone applications, the Poisson distribution was gradually applied to other engineering problems. The following example adapted from Grant[7] shows an application to the occurrence of excessive rainfall:

Example 2

RAINSTORMS

Number of storms per station per year	*Observed number of occurrences*	*Total storms*	*Probability*	*Theoretical number of occurrences*
0	102	0	0.301	99.3
1	114	114	0.361	119.1
2	74	148	0.217	71.6
3	28	84	0.087	28.7
4	10	40	0.026	8.6
5	2	10	0.006	2.0
>5	0	0	0.002	0.7
TOTAL	330	396	1.000	330.0

$$m = \frac{396}{330} = 1.20 \qquad e^{-m} = 0.301 \qquad 330\ e^{-m} = 99.3$$

The first published examples of the Poisson distribution as

[7] Grant, Eugene L., "Rainfall Intensities and Frequencies," *Transactions*, American Society of Civil Engineers, vol. 103, 1938, pp. 384-388.

applied to traffic data are presented by Adams.[8] The following is one of Adams' examples:

Example 3

RATE OF ARRIVAL (Vere St.)
(Number of vehicles arriving per 10 second interval)

No. vehicles per 10 sec. period x	*Observed frequency*	*Total vehicles*	*Probability* $P(x)$	*Theoretical frequency*
0	94	0	0.539	97.0
1	63	63	0.333	59.9
2	21	42	0.103	18.5
3	2	6	0.021	3.8
>3	0	0	0.004	0.8
TOTAL	180	111	1.000	180.0

$$m = \frac{\text{total vehicles}}{\text{total periods}} = \frac{111}{180} = 0.617;\ e^{-.617} = .539$$

$$P(x) = \frac{(.617)^x}{x!} e^{-.617} = \frac{(.539)\,(.617)^x}{x!}$$

Calculated frequency $= 180\ P(x)$

Note: Since there were 111 vehicles in 180 ten-second periods, the hourly volume was 222.

Testing Goodness of Fit (χ^2 Test)

In each of the foregoing examples it has been postulated that a Poisson distribution having a parameter m whose value has been computed from the observed data describes the population that has been sampled. The observed distribution constitutes this sample. By inspection there is apparent agreement between the observed distribution (sample) and the theoretical distribution. The inference is then made that the postulated theoretical (Poisson) distribution is in fact the true population

[8] Adams, *op. cit.*

distribution. This inference is based, however, solely on inspection; a more rigorous basis for reaching such a conclusion is desired. One of several statistical tests of significance may be used for this purpose; the chi square (χ^2) test is appropriate to the present application. This test, which is described in Appendix G and illustrated below, provides for one of two decisions:

1. It is not very likely that the true distribution (of which the observed data constitute a sample) is in fact identical with the postulated distribution.
2. The true distribution (of which the observed data constitute a sample) could be identical with the postulated distribution.

It can be seen that either decision can be erroneously made. Decision 1 can be wrong if in fact the postulated distribution is the true distribution. On the other hand, Decision 2 can be wrong if the true distribution is in fact different from the postulated distribution. Statistical tests of significance allow for specifying the probability of making either of these types of error. Usually the probability of making the first type of error (incorrectly rejecting the postulated distribution when in fact it is identical with the true distribution) is specified and no statement is made with regard to the second type of error. The specification of the first type of error is expressed as a "significance level." Common significance levels are 0.01, 0.05, and 0.10. Thus, when a test is made at the 0.05 (5%) level, the engineer takes the chance that 5% of the rejected postulated distributions are in fact identical with the corresponding true distributions. For a complete discussion of the theory underlying this and other statistical tests of significance, the reader is referred to any standard text on statistics.

The technique of performing the χ^2 significance test is illustrated in Example 4. In the table of Example 4 the observed frequency of wrong connections is shown along with the postulated theoretical frequency distribution (computed in Example 1). The problem is to decide whether the observed data can be construed as a sample coming from the postulated

theoretical distribution. The computing techniques described in Appendix G yield a chi square value of 7.6 with 11 degrees of freedom. (It should be noted in the table of Example 4 that the frequencies of 3 or fewer wrong connections have been combined as have the frequencies of 15 or greater. This is done to meet the requirement of the χ^2 test that the theoretical frequency be at least 5 in any group. Thus, this combination results in 13 groups, with $13 - 2 = 11$ degrees of freedom.)

Example 4

χ^2 TEST OF CONNECTIONS TO WRONG NUMBER (Data from Example 1)

Number of wrong connections per period x	*Observed frequency* f	*Postulated theoretical frequency* F	$\frac{f^2}{F}$
$\leqq 3$	6	6.7	5.4
4	11	10.4	11.6
5	14	18.2	10.8
6	22	26.4	18.3
7	43	33.1	55.9
8	31	36.0	26.7
9	40	35.2	45.5
10	35	30.7	39.9
11	20	24.3	16.5
12	18	17.9	18.1
13	12	12.0	12.0
14	7	7.5	6.5
$\geqq 15$	8	8.6	7.4
TOTAL	267	267.0	274.6

$$\chi^2 = 274.6 - 267.0 = 7.6$$

From statistical tables or from Figure 3 the value of χ^2 at the 0.05 (5%) significance level for $v = 11$ is found to be 19.7.*

* In using tables of χ^2 care should be exercised to note the manner in which the table is entered with the significance level. If the table is so constructed that for a given number of degrees of freedom the value of χ^2 increases with decreasing percentiles (probabilities), the table is entered with the percentile corresponding to the *significance level*. If the table is such that the value of χ^2 increases with increasing percentiles, the table is entered with the *significance level subtracted from 1*.

This is known as the critical value for the significance test at the 5% level. Since the computed value of χ^2 (7.6) is less than the critical value, decision 2 is accepted. (The acceptance of decision 2 is rigorously stated, "There is no evidence to indicate that the true theoretical distribution differs from the postulated distribution.") Had the computed value of χ^2 exceeded the critical value, decision 1 would have been accepted.

Once the decision has been made that a particular postulated theoretical distribution appropriately represents the population from which the observed data came, the theoretical distribution can be used in place of the observed distribution for engineering analysis and action, since it will be free from the random variations present in the observed distribution.

Interpretation of Examples 1 to 3

When the engineer gathers field data he is usually interested in predicting performance for design or administrative purposes. Fitting a theoretical curve to empirical data is only one step in the overall objective. The Poisson distribution is of value only if it permits useful conclusions to be drawn. Below are some conclusions which can be drawn from each of examples 1 to 3.

Example 1:

a. Wrong connections follow a Poisson distribution with the parameter $m = 8.742$.
b. Probabilities computed from the Poisson distribution may be used to predict the occurrence of wrong connections and thus may serve as a basis for future design considerations. Future design might, for instance, provide additional facilities to compensate for the capacity consumed by the wrong connections. A possible alternative might be the provision of some method of preventing wrong connections.
c. The effectiveness of engineering changes may be evaluated in terms of a significant change in Poisson parameter.

Example 2:

a. The occurrence of rainstorms in the population considered follows the Poisson distribution with the parameter $m = 1.20$.

b. The probabilities may serve as a basis for future design. Consider for instance the following:

Let D_1 = the damage caused if one storm occurs during a year,

D_2 = the damage caused if two storms occur during a year,

. . . .

D_n = the damage caused if n storms occur during a year,

$P(1)$ = the probability of one storm during a year,

$P(2)$ = the probability of two storms during a year,

. . . .

$P(n)$ = the probability of n storms during a year.

Then the expected or average damage per year will be

$$D_{av.} = D_1\,P(1) + D_2\,P(2) + \ldots\ldots\ldots + D_n\,P(n).$$

Example 3:

a. At the location considered and during the time of day and day of the week considered the arrival of vehicles follows the Poisson distribution with the parameter $m = 0.617$ for 10-second intervals.

b. The probabilities may be used for design of traffic control measures. (Examples appearing later in this paper indicate some specific methods of applying probabilities to traffic design problems.)

Traffic Applications of the Poisson Distribution

Prediction of Arrivals from Hourly Volume

For conditions in which the Poisson distribution applies—i.e., under conditions of "free flow"—it is possible to compute the probability of 0, 1, 2, ... , k vehicles arriving per time interval of t seconds provided the hourly volume, V, is known:

t = length of time interval in seconds

V = hourly volume

n = number of intervals (per hour)

$$= \frac{3600}{t}$$

m = average number of vehicles per interval

$$= \frac{V}{\dfrac{3600}{t}} = \frac{Vt}{3600}$$

Then the probability, $P(x)$, that x vehicles will arrive during any interval is:

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{1}{x!}\left(\frac{Vt}{3600}\right)^x e^{-\frac{Vt}{3600}}$$

The hourly frequency, F_x, of intervals containing x vehicles is:

$$F_x = n\,P(x) = \left(\frac{3600}{t}\right)\frac{1}{x!}\left(\frac{Vt}{3600}\right)^x e^{-\frac{Vt}{3600}}$$

If the period under consideration is different from an hour, the 3600 in the first parentheses would be replaced by the appropriate length of time in seconds. The value V, however, would still be the hourly volume. Examples 5 and 6 illustrate this.

Example 5

Prediction of Arrivals (Low Volume)

V = Hourly volume = 37

t = Length of interval = 30 seconds

$$m = \frac{Vt}{3600} = \frac{37 \cdot 30}{3600} = 0.308$$

$$e^{-m} = 0.735$$

$$n = \frac{3600}{t} = \frac{3600}{30} = 120$$

$$F_0 = n\left(\frac{1}{0!}\right) m^0 e^{-m} = (120)\ (1)\ (1)\ (0.735) = 88.2$$

$$F_1 = F_0\left(\frac{m}{1}\right) = (88.2)\ (0.308) = 27.2$$

$$F_2 = F_1\left(\frac{m}{2}\right) = \frac{(27.2)\ (0.308)}{2} = 4.2$$

$$F_{>2} = n - (F_0 + F_1 + F_2) = 120 - (88.2 + 27.2 + 4.2) = 0.4$$

Note: The method of computing F_1 from F_0 and F_2 from F_1 is discussed in section on Methods of Computation.

The following table compares the predicted frequencies with frequencies observed in a field study (Elliot Avenue Westbound).[9] The predicted frequency in this example corresponds to the theoretical frequency of the previous examples.

Number of vehicles arriving per interval x	*Observed*[9] *frequency* f_x	*Predicted frequency* F_x
0	88	88.2
1	27	27.2
2	5	4.2
$\geqq 3$	0	0.4
	120	120.0

Example 6

PREDICTION OF ARRIVALS (Moderate Volume)
(30-Second Intervals for 25 *Minutes*)

$V = 418$

$t = 30$ seconds

$$m = \frac{418 \cdot 30}{3600} = 3.48$$

$$n = \frac{25 \cdot 60}{30} = \frac{1500}{30} = 50 \text{ (25 minutes, 60 seconds per minute)}$$

$e^{-m} = 0.0308$

$F_o = n\, e^{-m} = 50 \cdot 0.308 = 1.54$

etc.

In the following table the various predicted frequencies are compared with frequencies observed in the field (Redondo Beach Blvd. Eastbound).

[9] Based on field data supplied through the courtesy of the Traffic and Lighting Division, Los Angeles County Road Department.

Number of vehicles arriving per interval x	*Observed frequency* f_x	*Predicted frequency* F_x
0	0	1.5
1	9	5.4
2	6	9.3
3	9	10.8
4	11	9.4
5	9	6.6
6	5	3.8
7	1	1.9
$\geqq 8$	0	1.3
	50	50.0

The following table shows the χ^2 analysis of the data above:

Number vehicles arriving per interval x	*Observed frequency* f_x	*Predicted (theoretical) frequency* F_x	$\frac{f_x^2}{F_x}$
$\leqq 1$	9	6.9	11.7
2	6	9.3	3.9
3	9	10.8	7.5
4	11	9.4	12.9
5	9	6.6	12.3
$\geqq 6$	6	7.0	5.1
	50	50.0	53.4

$$\sum \frac{f_x^2}{F_x} - n = 53.4 - 50.0 = 3.4$$

$$v = 6 - 2 = 4$$

$\chi^2_{.05} = 9.49$ (The symbol $\chi^2_{.05}$ is used to denote the critical value of χ^2 at the 0.05 (5%) significance level.

The agreement between the predicted and observed data is acceptable at the 5% significance level.

Testing Goodness of Fit with a Poisson Population

In some locations traffic parameters will vary with time

throughout the day (or week or year) while at other locations the variability may appear to be entirely random. One way in which the traffic engineer may test the randomness of his data is to assume that it follows the Poisson distribution, and then to test this assumption by the χ^2 test. If the χ^2 test fails, there is strong evidence that the data are non-random. If the results of the χ^2 test are acceptable, the data may be assumed to be random for the purposes of the traffic engineer. It should be noted, however, that an acceptable χ^2 test is less conclusive than one which fails. The rigorous interpretation of an acceptable test should be, "There is no evidence to indicate non-randomness of the data."

Example 7 is an analysis of right turns during 300 3-minute intervals distributed throughout various hours of the day and various days of the week. While casual inspection of the observed data might lead one to suspect randomness, the test shows strong evidence of non-randomness.

Example 7

RIGHT TURNS (Westwood at Pico)

Number of cars making right turns per 3-min. interval	*Observed frequency* f		*Theoretical frequency* F		$\frac{f^2}{F}$
0	14		6.1		32.1
1	30		23.7		38.0
2	36		46.2		28.1
3	68		59.9		77.2
4	43		58.3		31.7
5	43		45.4		40.7
6	30		29.4		30.6
7	14		16.4		12.0
8	10		8.0		12.5
9	6		3.4		
10	4		1.3		
11	1	12	0.6	6.6	21.8
12	1		0.3		
$\geqq 13$	0		1.0		
	300		300.0		324.7

(The brackets indicate grouping for the χ^2 test.)

$$m = \frac{1168}{300} = 3.893 \qquad e^{-m} = 0.0203$$

$$\sum \frac{f^2}{F} - n = 324.7 - 300.0 = 24.7$$

$$v = 10 - 2 = 8$$

$$\chi^2_{.05} = 15.5$$

Since $24.7 > 15.5$, the fit is *not* acceptable at the 5% significance level.

Some traffic phenomena may be random when observed for an interval of one length but non-random when observed with an interval of a different length. Example 8a shows arrival rates for 10-second intervals. Here the failure of the χ^2 test (non-acceptability of fit) indicates non-randomness. Example 8b considers the same arrivals with 30-second observation intervals; the assumption of randomness acceptable at the 5% level.

Example 8a

ARRIVAL RATE—10-SECOND INTERVALS (Durfee Avenue, Northbound)

Number of cars per interval	*Observed frequency*[10] f	*Theoretical frequency* F	$\frac{f^2}{F}$
0	139	129.6	149.1
1	128	132.4	123.7
2	55	67.7	44.7
3	25	23.1	27.1
4	10 }	5.9 }	
5	3 } 13	1.2 } 7.2	23.5
≥ 6	0 }	0.1 }	
	360	360.0	368.1

(As before, the brackets indicate grouping for the χ^2 test)

$$m = \frac{368}{360} = 1.022 \qquad e^{-m} = 0.360$$

[10] See footnote 9.

$$\sum \frac{f^2}{F} - n = 368.1 - 360.0 = 8.1$$

$$v = 5 - 2 = 3$$

$$\chi^2_{.05} = 7.81$$

The fit is, therefore, *not* acceptable at the 5% level.

Example 8b

ARRIVAL RATE—30-SECOND INTERVALS (Durfee Avenue, Northbound)

Number of cars per interval	*Observed frequency*[10] f	*Theoretical frequency* F	$\frac{f^2}{F}$
0	9	5.6	14.5
1	16	17.2	14.9
2	30	26.3	34.2
3	22	26.9	18.0
4	19	20.6	17.5
5	10	12.6	7.9
6	3	6.5	
7	7 } 14	2.8 } 10.8	18.1
8	3	1.1	
≧ 9	1	0.4	
	120	120.0	125.1

$$m = \frac{368}{120} = 3.067 \qquad e^{-m} = 0.047$$

$$\sum \frac{f^2}{F} - n = 125.1 - 120.0 = 5.1$$

$$v = 7 - 2 = 5$$

$$\chi^2_{.05} = 11.1$$

This fit *is* acceptable at the 5% significance level.

Probability of Finding a Vacant Parking Space

Certain analyses of parking may be treated by the Poisson distribution. Example 9 shows the results of a study of 4 block

[10] See footnote 9.

faces containing 48 one-hour parking spaces. (Loading zones and short-time parking spaces were not included). Observations were taken during the hours of 2 P.M. to 4 P.M. on the days Monday through Friday. The observation period was divided into 5-minute intervals. During each 5-minute interval one observation was made; the exact time within each interval was randomized. Each observation consisted of an instantaneous count of the number of vacant spaces. Since the χ^2 test shows good agreement between the observed and theoretical data, there is no ground to doubt that the distribution of vacant parking spaces follows the Poisson distribution. Assuming the Poisson distribution to hold, it is possible to compute the probability of finding a vacant one-hour parking space within these four block faces.

Example 9

VACANT PARKING SPACES (Westwood Blvd.)

Number of vacant 1-hour parking spaces per observation	*Observed frequency* f	*Theoretical frequency* F	$\frac{f^2}{F}$
0	29	25.1	33.5
1	42	39.3	44.9
2	21	30.1	14.7
3	16	16.4	15.6
4	7	6.4	
5	2	2.0	
6	3 } 12	0.5 } 9.1	15.8
$\geqq 7$	0	0.2	
	120	120.0	124.5

$$m = \frac{188}{120} = 1.567 \qquad e^{-m} = 0.209$$

$$\sum \frac{f^2}{F} - n = 124.5 - 120.0 = 4.5$$

$$v = 5 - 2 = 3$$

$$\chi^2_{.05} = 7.81$$

The probability of finding one or more vacant parking spaces, $P\ (\geqq 1)$, is

$$P\ (\geqq 1) = 1 - P(O) = 1 - e^{-m} = 1 - 0.209 = 0.791$$

Analysis and Comparison of Accident Data

Several writers have discussed the use of the Poisson distribution in accident analysis. The example which follows is taken from an unpublished paper by Belmont.[11]

Ninety-three road sections each having average traffic less than 8,000 vehicles per day and average usage of less than 40,000 vehicle miles per day were combined to give 45 composite roads each carrying 35,000 to 40,000 vehicle-miles per day. The accident records of these roads were then compared using the Poisson distribution as follows:

Example 10

SINGLE-CAR ACCIDENTS

Number of single-car accidents 1950 x	*Number of roads observed* f_x		*Theoretical number of roads* F_x		$\frac{f_x^2}{F_x}$
0	18		14.5		22.3
1	14		16.4		12.0
2	7	13	9.3	14.1	12.0
3	4		3.5		
4	1		1.0		
5	0		0.2		
6	0		0.04		
7	1		0.007		
$\geqq 8$	0		0.001		
	45		45		46.3

$$m = \frac{51}{45} = 1.133 \qquad e^{-m} = 0.322 \qquad 45e^{-m} = 14.5$$

[11] Belmont, D. M., *The Use of the Poisson Distribution in the Study of Motor Vehicle Accidents,* Technical Memorandum No. B-7, Institute of Transportation and Traffic Engineering, University of California, June 20, 1952.

$$\sum \frac{f_x^2}{F_x} = 46.3 - 45.0 = 1.3 \qquad v = 3 - 2 = 1$$

$$\chi^2_{.05} = 3.84$$

Although the fit is acceptable at the 5% level, one might be suspicious of the road section having 7 accidents when the probability of this occurrence is $\frac{0.007}{45}$ or approximately 2×10^{-4}. Eliminating this one road results in the following:

Example 10b

SINGLE-CAR ACCIDENTS

Number of single-car accidents 1950 x	*Number of roads observed* f_x	*Theoretical number of roads* F_x	$\frac{f^2}{F_x}$
0	18	16.2	20.0
1	14	16.2	12.1
2	7 }	8.1 }	
3	4 } 12	2.7 } 11.6	12.4
4	1 }	0.7 }	
$\geqq 5$	0 }	0.1 }	
	44	44.0	44.5

$$m = \frac{44}{44} = 1.000 \qquad e^{-m} = 0.368 \qquad 44e^{-m} \doteq 16.2$$

$$\sum \frac{f_x^2}{F_x} = 44.5 - 44.0 = 0.5 \qquad v = 3 - 2 = 1$$

$$\chi^2_{.05} = 3.84$$

Since there is no evidence (at the 5% level of significance) that there is a difference between the observed and the theoretical data, one may reasonably assume that each of these 44 roads has the same accident potential. Thus, among road sections from this group, any section is as likely as any other to have from zero to four accidents per year; a section which has four

accidents is no more dangerous than one having no accidents, the occurrences being simply the operation of chance.

When only two roads or intersections are to be compared as to accident potential, the following technique[12] may be used:

a. Compute u,

$$\text{where } u = \frac{x_1 - x_2 - 1}{\sqrt{x_1 + x_2}}$$

x_1 = larger of observed values

x_2 = smaller of observed values

b. Compare the resulting value of u with the appropriate critical value selected from table. If the computed value of u is less than the critical value, the two roads or intersections may be considered to have the same accident potential.

Significance level	*Critical value of u*
0.01	2.58
0.05	1.96
0.10	1.64

Example 11

COMPARISON OF ACCIDENTS AT TWO INTERSECTIONS

Intersection 1: 6 accidents during year considered
2: 8 accidents during year considered

$$x_1 = 8 \qquad x_2 = 6$$

$$u = \frac{8 - 6 - 1}{\sqrt{8 + 6}} = 0.267$$

Since the computed value of u in Example 11 is less than the 5% significance level, there is no evidence of a difference in accident potential between the two intersections.

[12] Hald, A., *Statistical Theory with Engineering Applications,* John Wiley and Sons, Inc., 1952, pp. 725-726.

Cumulative Poisson Distribution

In the previous discussion, the probability of exactly 0, 1, 2, . . . items per trial (or time interval) has been computed. In many problems it is desirable to compute such values as the probability that the number of items (vehicles arriving) per trial is:

a. k or less
b. greater than k
c. less than k
d. k or greater

These probabilities involve the cumulative Poisson distribution and may be expressed as follows:

$$P(x \leq k) = \text{Probability that } x \leq k$$

$$= \sum_{x=0}^{k} P(x) = \sum_{x=0}^{k} \frac{m^x}{x!} e^{-m}$$

$$P(x > k) = 1 - P(x \leq k) = 1 - \sum_{x=0}^{k} \frac{m^x}{x!} e^{-m}$$

$$P(x < k) = P(x \leq k-1) = \sum_{x=0}^{k-1} \frac{m^x}{x!} e^{-m}$$

$$P(x \geq k) = 1 - P(x < k) = 1 - \sum_{x=0}^{k-1} \frac{m^x}{x!} e^{-m}$$

The following examples illustrate the use of the cumulative Poisson distribution.

Example 12

Left-turn Control

This example is patterned after one given by Adams.[13]

A single intersection is to be controlled by a fixed-time signal having a cycle of 55 seconds. From one of the legs there is a left-turning movement amounting to 175 vehicles per hour. The

[13] Adams, *op. cit.*

layout of the intersection is such that two left-turning vehicles per cycle can be satisfactorily handled without difficulty, whereas three or more left-turning vehicles per cycle cause delays to other traffic. In what percent of the cycles would such delays occur?

$$m = \text{avg. left turns per cycle} = \frac{55 \times 175}{3600} = 2.67$$

$$P(x \geq 3) = 1 - \sum_{x=0}^{2} \frac{(2.67)^x}{x!} e^{-2.67}$$

$$= 1 - [0.069 + 0.185 + 0.247]$$

$$= 1 - 0.501 = 0.499$$

Answer: 49.9%

If a special left-turn phase is provided, in what percent of the cycles will this special feature be unnecessary by virtue of the fact that there is no left-turning vehicle?

$$P(0) = e^{-2.67} = 0.069$$

Answer: 6.9%

Example 13

Left-turn Storage Pockets

The California Division of Highways has applied the cumulative Poisson distribution to the design of left-turn storage pockets. The pockets are designed in such a manner that the number of cars desiring to make a left-turn during any signal cycle will exceed the pocket capacity only 4% of the time—i.e., the probability of greater than a given number of cars is 0.04. This may be expressed:

$$0.04 = 1 - \sum_{x=0}^{k} \frac{1}{x!} \left(\frac{Lt}{3600}\right)^x e^{-\frac{Lt}{3600}}$$

Where: L = number of left-turning cars per hour
t = length of signal cycle in seconds
k = pocket capacity (number of cars)

This relationship may be solved for k by accumulating terms in the right-hand member until the equation is satisfied.* A more direct method of computation, however, is to use tables or charts showing the various values for the cumulative Poisson distribution. Such a chart is included in the section on Methods of Computation.

The values obtained are as follows:

Required Storage (Number of Vehicles)

Peak hour left-turn movements	*60 second cycle*	*120 second cycle*
100	4	7
200	7	12
300	9	16
400	12	20
500	14	24
600	16	28
700	18	32
800	20	36

The Exponential Distribution

The previous discussions of arrival rate, etc., have dealt with the distribution of discrete events (arrivals of cars) within a given time interval. The distribution of gaps (time spacing) between vehicles is a continuous variable and is exponential in nature. This distribution may be derived from the Poisson distribution as follows:

As before, let:

$$m = \text{average number of cars per time interval}$$

$$= \frac{t}{3600} V$$

* The equation will in general never be precisely satisfied because of discrete nature of the Poisson distribution. There will be two values of k which nearly satisfy the relationship—one too large and one too small. It is then a matter of engineering judgment which to use.

Where V = hourly volume

t = length of interval in seconds

$P(o)$ = Probability of zero vehicles during an interval t

$$P(o) = \frac{m^o e^{-m}}{O!} = e^{-m} = e^{-\frac{Vt}{3600}} \qquad \text{(A)}$$

If there are no vehicles in a particular interval of length t, then there will be a gap of *at least* t seconds between the last previous vehicle and the next vehicle. In other words, $P(o)$ is also the probability of a gap equal to or greater than t seconds. This may be expressed:

$$P(g \geq t) = e^{-\frac{Vt}{3600}} \qquad \text{(B)}$$

From this relationship it may be seen that (under conditions of random flow) the number of gaps greater than any given value will be distributed according to an exponential curve. This is known as an exponential distribution.

Example 14

Distribution of Gaps (Arroyo Seco Freeway)

Total gaps = 214

Total time = 1753 sec.

$$m = \frac{214}{1753}t = 0.122t$$

$$P(g \geq t) = e^{-0.122t}$$

$$G = (\text{Expected number of gaps} \geq t) = 214e^{-0.122t}$$

Figure 1 shows a comparison between the observed and theoretical results.[14]

[14] When using equation [B] it should be noted that $\frac{Vt}{3600}$, the number of vehicles arriving during an interval, and G, the number of gaps equal to or greater than some specified value t, are discrete variables. That is, the number of vehicles or the number of gaps can take on only integral values 0, 1, 2, ... etc. Probability, $P(x)$ and time, t, on the other hand, are continuous variables and can take on fractional values. This is illustrated in Figure 1, where the theoretical frequencies are represented by bars and the probability curve is represented by a continuous dashed curve.

Example 15

Safe Gaps at School Crossings

(This example is based on the report of a joint committee of the Institute of Traffic Engineers and the International Association of Chiefs of Police.[15])

In studying the natural gaps in traffic at school crossings the following assumptions may be made:

1. The walking speed of a child is 3.5 ft./sec.
2. There must be at least one opportunity of crossing per minute. (This implies a minimum of 60 opportunities per hour.)

On the basis of these assumptions it is desired to determine the critical volume for a street of a given width. The critical volume here refers to that volume above which special measures will be required for the safety of the child.

There are two approaches to this problem, and care must be exercised in the selection of the appropriate method. One approach is that of the number of gaps greater than the time t required for a child to cross the street. The expected number of gaps per hour will be V, and the probability (fraction) of gaps equal to or greater than t is as derived in equation [B]. Thus, the expected number of gaps per hour which are equal to or greater than t will be

$$Ve^{-\frac{Vt}{3600}}$$

The other approach is the number of t-second intervals per hour which are free of cars. The number of t-second intervals per hour is $3600/t$. The probability (fraction) of such intervals free of cars is as given in equation [A]. Thus, the expected number of t-second intervals per hour which are free of cars will be

$$\frac{3600}{t}e^{-\frac{Vt}{3600}}$$

[15] "Report on Warrants for Traffic Officers at School Intersections," *Proceedings,* Institute of Traffic Engineers, vol. 18, 1947, pp. 118-130.

When one considers the fact that a long gap may contain several t-second intervals during which there is opportunity to cross, this second approach seems more appropriate.

Introducing the second assumption (at least 60 opportunities to cross per hour),

$$\frac{3600}{t} e^{-\frac{V_c t}{3600}} = 60$$

where V_c denotes the critical volume.

From the first assumption (walking speed = 3.5 ft/sec),

$$t = \frac{D}{3.5}$$

where D is the width of the street.

Substituting:

$$\frac{3600 \cdot 3.5}{D} e^{-\frac{V_c D}{3600 \cdot 3.5}} = 60$$

$$e^{-\frac{V_c D}{12,600}} = \frac{60D}{12,600} = \frac{D}{210}$$

Taking the natural logarithm of both sides:

$$\frac{-V_c D}{12,600} = ln \frac{D}{210}$$

$$V_c = -\frac{12,600}{D} ln \frac{D}{210}$$

Making use of the relationship

$$ln\ x = \frac{\log x}{\log e} = \frac{\log x}{0.4343}$$

gives

$$V_c = -\frac{12{,}600}{D}\frac{1}{0.4343}\log\frac{D}{210}$$

or

$$V_c = -\frac{29{,}000}{D}(\log D - \log 210)$$

or

$$V_c = \frac{29{,}000}{D}(2.322 - \log D)$$

This equation has been adopted by the joint committee of the Institute of Traffic Engineers and the International Association of Chiefs of Police. Solution of the equation gives the following values

Width of street (ft.):	25	50	75
Critical volume (v.p.h.):	1072	361	173

Miscellaneous Techniques

Multiple Poisson Distribution

When one group of events follows the Poisson distribution with the parameter m_1 and another group of events follows the Poisson distribution with the parameter m_2, the population formed by combining these groups will follow a Poisson distribution with the parameter $(m_1 + m_2)$. (A formal mathematical proof of this relationship is given in section H of the Appendix.) From the case for combining 2 groups of events it follows that when k groups of events are combined the resulting Poisson distribution has the parameter $(m_1 + m_2 + m_3 + \ldots\ldots\ldots\ldots + m_k)$. One traffic application of this relationship is the mathematical description of the total arrivals at an intersection when the arrivals on each leg are known to follow the Poisson distribution. Example 16 illustrates this application.

Example 16a

ARRIVAL RATE—SOUTHBOUND (Durfee Ave.)[16]

Number of cars arriving per 30-sec. interval	*Observed frequency* f	*Theoretical frequency* F	$\frac{f^2}{F}$
0	18	15.0	21.6
1	32	31.2	32.8
2	28	32.4	24.2
3	20	22.5	17.8
4	13	11.7	14.4
5	7 } 9	4.9 } 7.2	11.3
6	0	1.6	
7	1	0.5	
8	1	0.1	
$\geqq 9$	0	0.1	
	120	120.0	122.1

(Brackets indicate grouping for χ^2 test.)

$$m_s = \frac{250}{120} = 2.08 \qquad e^{-m_s} = 0.125$$

$$\sum \frac{f^2}{F} - n = 122.1 - 120.0 = 2.1$$

$$v = 6 - 2 = 4$$

$$\chi^2_{.05} = 9.49$$

16 See footnote 9.

Example 16b

ARRIVAL RATE—NORTHBOUND (Durfee Ave.)

Number of cars arriving per 30-sec. interval	*Observed frequency* f	*Theoretical frequency* F	$\frac{f^2}{F}$
0	9	5.6	14.5
1	16	17.2	14.9
2	30	26.3	34.2
3	22	26.9	18.0
4	19	20.6	17.5
5	10	12.6	7.9
6	3	6.5	
7	7	2.8	
8	3	1.1	
$\geqq 9$	1	0.4	
6–$\geqq 9$ (grouped)	14	10.8	18.1
	120	120.0	125.1

(Brackets indicate grouping for the χ^2 test.)

$$m_n = \frac{368}{120} = 3.07 \qquad e^{-m_n} = 0.047$$

$$\sum \frac{f^2}{F} - n = 125.1 - 120.0 = 5.1$$

$$v = 7 - 2 = 5$$

$$\chi^2_{.05} = 11.1$$

Example 16c

Arrival Rate—Northbound and Southbound Combined
(Durfee Ave.)

Number of cars arriving per 30-sec. interval	*Observed frequency* f		*Theoretical frequency* F		$\frac{f^2}{F}$
0	1		0.7		
1	4	14	3.6	13.5	14.5
2	9		9.2		
3	19		15.9		22.7
4	20		20.4		19.6
5	16		21.0		12.2
6	19		18.0		20.1
7	10		13.2		7.6
8	13		8.5		19.9
9	4		4.9		
10	1		2.5		
11	3	9	1.2	9.5	8.5
$\geqq$ 12	1		0.9		
	120		120.0		125.1

$$m_{ns} = \frac{618}{120} = 5.15 \qquad m_n + m_s = 3.07 + 2.08 = 5.15$$

$$e^{-m_{ns}} = 0.006$$

$$\sum \frac{f^2}{F} - n = 125.1 - 120.0 = 5.1$$

$$v = 8 - 2 = 6$$

$$\chi^2_{.05} = 15.5$$

Empirical Formulas Based on the Poisson Distribution

Because of differences between theoretical and physical conditions it is frequently necessary to develop empirical formulas to adequately represent the true behavior of traffic. Where

traffic arrives in a random manner, the Poisson distribution may serve as a building block for such formulas.

Consider the delays at urban stop signs reported by Raff.[17] If a gap of t seconds is required for crossing the main street, and if the side street cars arrive at random, the percent of side street cars which cross *without delay* will be equal to the *percent* of gaps in the main street traffic which are t or greater. From the equation B this is:

$$100e^{-\frac{Vt}{3600}}$$

The percent of side street cars delayed will, therefore, be

$$100\left(1-e^{-\frac{Vt}{3600}}\right)$$

Raff has found empirically that, because of sluggishness of the side traffic, the delays should be much greater than the above value at high volumes on the side street. He has developed the following empirical formula to represent this situation.

$$P_{\%} = 100\left\{1 - \frac{e^{-2.5V_s/3600}\, e^{-2Vt/3600}}{1 - e^{-2.5V_s/3600}\left(1 - e^{-Vt/3600}\right)}\right\}$$

where V_s is the side street volume.
When $V_s = 0$, this reduces to the original relationship.

Combination of Poisson Distribution and Other Relationships

Several writers have used the Poisson distribution in combination with other relationships to form theoretical expressions of various traffic problems. Although these techniques are beyond the scope of the present discussion, the reader may wish to familiarize himself with some of the papers on the subject. The following partial list of references is suggested as a starting point in the study of these advanced techniques.

Garwood, F., "An Application of the Theory of Probability to the Operation of Vehicular-Controlled Traffic Signals," *Journal* of

[17] Raff, Morton S., *A Volume Warrant for Urban Stop Signs,* The Eno Foundation for Highway Traffic Control, 1950.

The Royal Statistical Society, Supplement, vol. 7, no. 1, 1940, pp. 65-77.

Raff, Morton S., "The Distribution of Blocks in an Uncongested Stream of Automobile Traffic," *Journal* of The American Statistical Association, vol. 46, no. 253, March 1951, pp. 114-123.

Kendall, David G., "Some Problems in the Theory of Queues," *Journal* of the Royal Statistical Society, Series B, vol. 13, no. 2, 1951, pp. 151-173.

Tanner, J. C., "The Delay to Pedestrians Crossing a Road," *Biometrika,* vol. 38, parts 3 and 4, Dec. 1951, pp. 383-392.

Pearcy, T., "Delays in Landing of Air Traffic," *Journal* of The Royal Aeronautical Society, vol. 52, Dec. 1948, pp. 799-812.

Statistics of the Poisson Distribution

In section F of the Appendix the mean, variance, and standard deviation of the Poisson distribution are derived. The results are:

Mean: $\mu_x = m$

Variance: $\sigma_x^2 = m$

Standard Deviation: $\sigma_x = \sqrt{m}$

For large values of m the section of the probability curve between $\mu_x - \sigma_x$ and $\mu_x + \sigma_x$ is approximately 68% of the total area; the section between $\mu_x - 2\sigma_x$ and $\mu_x + 2\sigma_x$ is approximately 95% of the total area.

Methods of Computation

Review of Procedure

The following is a review of the procedure of computing probabilities from the Poisson distribution:

1. Determine the parameter m. This parameter is the average number of occurrences. It may be determined from observed or assumed data. A trial may consist of an instantaneous observation, counting events during a time interval, counting events in a unit area, etc.

$$m = \frac{\text{Total number of events observed}}{\text{Total number of trials or time intervals, etc.}}$$

2. Once m has been determined, the probability of x events occurring at any trial (during any time interval) is computed from the formula

$$P(x) = \frac{m^x e^{-m}}{x!}$$

$$\text{where } x! = x(x-1)\ (x-2) \cdots 3 \cdot 2 \cdot 1$$

Table 1 lists several sources of tables of e^m, e^{-m}, $P(x)$, etc. Figure 2 is a chart of the cumulative Poisson distribution, and may be used for rapid calculations where accuracy is not important. A method of making slide rule calculations is indicated below.

Chart of the Cumulative Poisson Distribution

Figure 2 is a modification of charts by Thorndike[18] and Working.[19] The probability $P(x \leqq c)$ is plotted against m with c as a parameter. The probability of x equal to or less than c is simply read from the curve. To obtain $P(x = c)$, the probability of exactly c, values are read from the chart for $x = c$ and for $x = (c - 1)$. The following relationship is then used:

$$P(x = c) = P(x \leqq c) - P(x \leqq c - 1)$$

Slide Rule Calculations

The value of e^x can be found by means of a log-log-duplex slide rule; e^{-x} is, of course, the reciprocal of e^x. Having obtained e^{-x} by slide rule or from tables, the individual terms of the Poisson distribution may be obtained by means of the following relationship for $P(x + 1)$ which is particularly adapted to slide rule calculations.

$$P(x) = \frac{m^x e^{-m}}{x!}$$

18 Thorndike, *op. cit.*

19 Working, Holbrook, *A Guide to Utilization of the Binomial and Poisson Distributions in Industrial Quality Control,* Stanford University Press, 1943.

$$P(x+1) = \frac{m^{x+1}e^{-m}}{(x+1)!} = \frac{m^x e^{-m}\, m}{x!\,(x+1)} = \frac{m}{x+1} P(x)$$

Thus, it follows that:

$$P(0) = e^{-m}$$

$$P(1) = \frac{m}{1} P(0)$$

$$P(2) = \frac{m}{2} P(1)$$

$$P(3) = \frac{m}{3} P(2)$$

etc.

Table 1

AIDS TO COMPUTATION OF POISSON DISTRIBUTION

Type of Aid and Scope	*Source*
e^x, e^{-x} Tables x from 0.00 to 5.50 by 0.01 steps. x from 5.0 to 10.0 by 0.1 steps	Table of Exponentials Handbook of Chemistry and Physics Chemical Rubber Publishing Company, Cleveland Various editions
$\frac{a^x\, e^{-a}}{x!}$ Tables Individual and Cumulative a from 0.001 to 100.0 by various steps	Molina, E. C., "Poissons Exponential Binomial Limit," Van Nostrand, 1945
$\frac{e^{-\epsilon}\, \epsilon^j}{j!}$ Tables ϵ from 0.1 to 1.0 by steps of 0.1. ϵ from 1.0 to 20.0 by steps of 1.0 Individual and cumulative	Appendices VI and VII Fry, T. C., "Probability and Its Engineering Uses, Van Nostrand, 1928

Table 1 *(Continued)*

Type of Aid and Scope	*Source*
$\frac{m^x e^{-x}}{x!}$ Cumulative Tables m from 0.1 to 15.0 by steps of 0.1	Appendix VI Greenshields, Bruce D.; Shapiro, Donald; Ericksen, Elroy L.; "Traffic Performance at Urban Street Intersections," Technical Report No. 1, Yale Bureau of Highway Traffic, 1947.
Chart Probability that no. cars arriving in a given time will be x or more (or less than x)	Figure 3 Adams, William F., "Road Traffic Considered as a Random Series," Institution of Civil Engineers Journal, Nov., 1936, pp. 121-130+.
Chart Cumulative $P(x \geqq c)$	Figure 5 Thorndike, Frances, "Applications of Poissons Probability Summation," Bell System Technical Journal, vol. 5, No. 4, Oct., 1926, pp. 604-624.
Slide Rule e^{-x} x from 0.001 to 10.0 on 3 or 4 10-inch scales plus normal slide rule features	The Fredrick Post Company Versalog Slide Rule x from 0.001 to 10.0 on 4 scales. Keuffel and Esser Company Log-Log-Duplex Trig Slide Rule Log-Log-Duplex Decitrig Slide Rule x from 0.01 to 10.0 on 3 scales.
e^x and e^{-x} Tables x from 0.0 to 0.1 by steps of 0.0001; from 0.1 to 3.0 by steps of 0.001; from 3.0 to 6.3 by steps of 0.01; from 6.3 to 10 by steps of 0.1	Hayashi, Keiichi, "Fünfstellige Tafeln der Kreis- und Hyperbelfunktionen sowie der Funktionen e^x und e^{-x}," Berlin, 1944, Walter de Gruyter & Co.

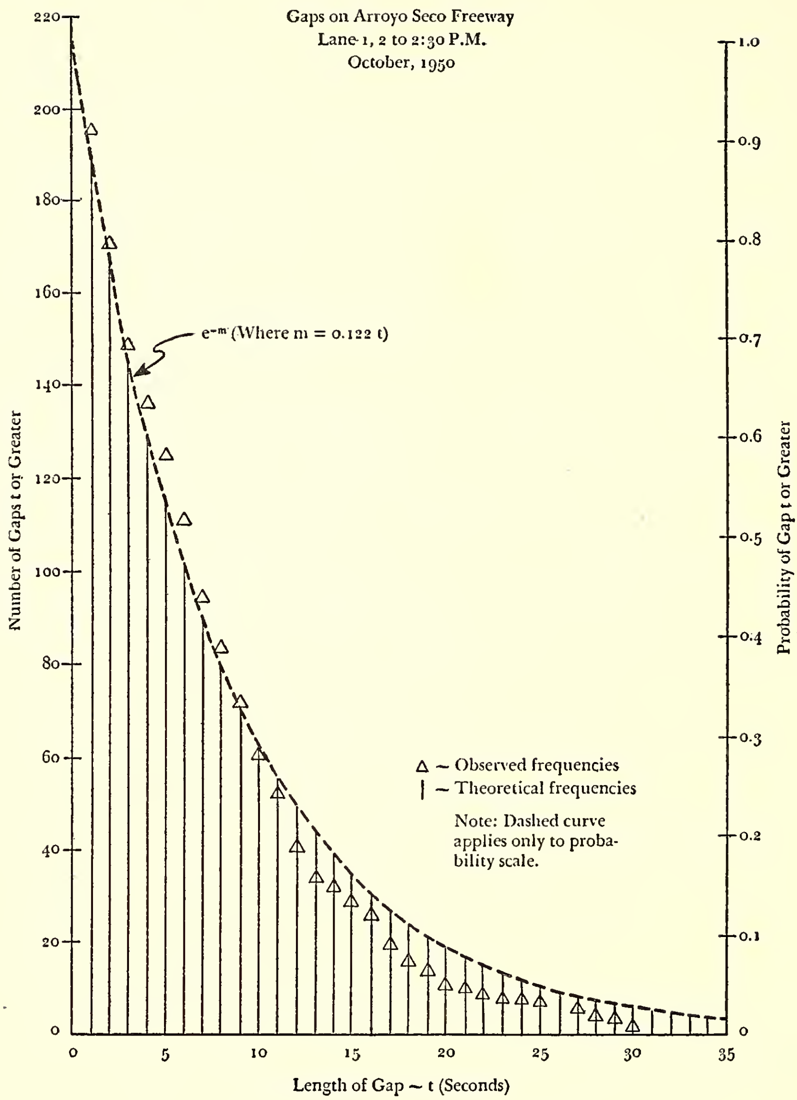
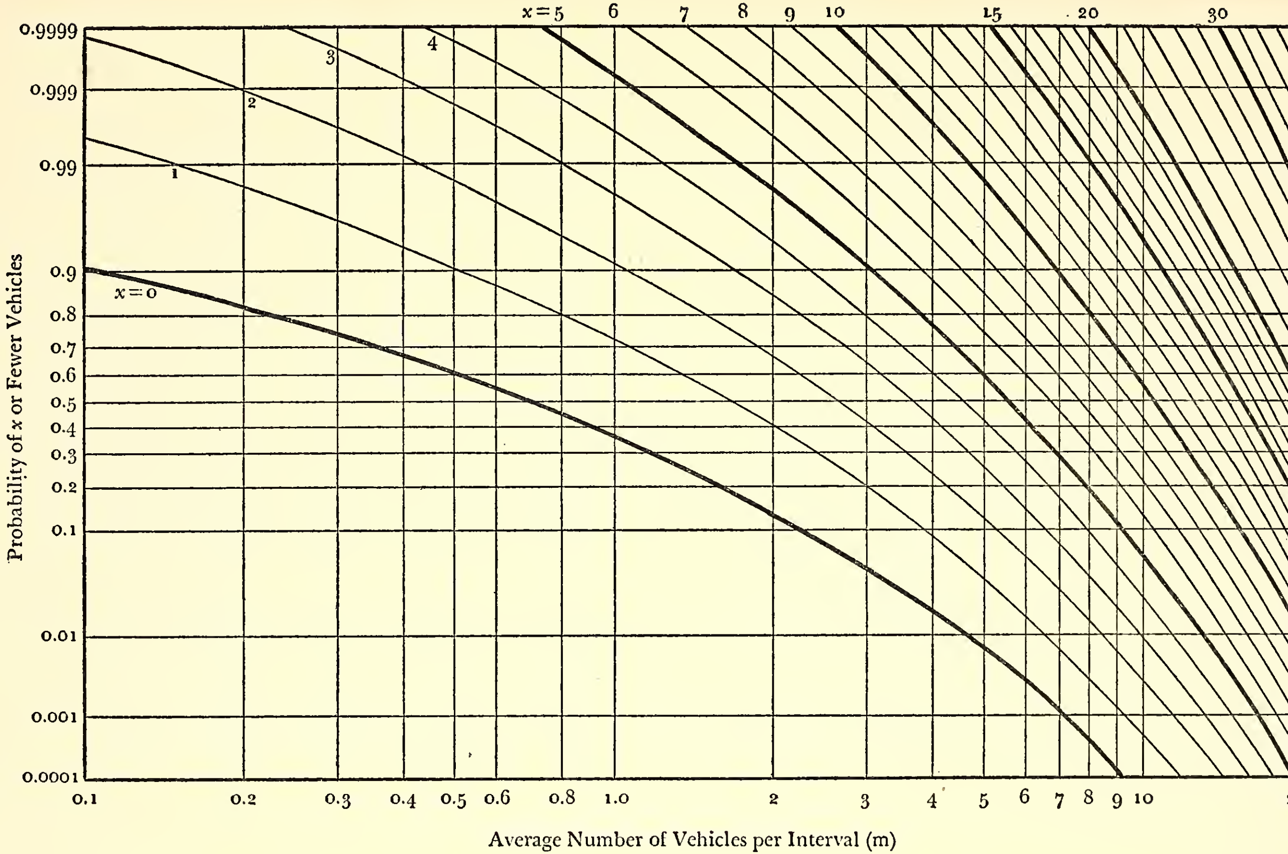

Gaps on Arroyo Seco Freeway
Lane 1, 2 to 2:30 P.M.
October, 1950
220
200
180
160
140
120
100
80
60
40
20
0
Number of Gaps t or Greater
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Probability of Gap t or Greater
e^-m (Where m = 0.122 t)
Δ ~ Observed frequencies
| ~ Theoretical frequencies
Note: Dashed curve applies only to probability scale.
0
5
10
15
20
25
30
35
Length of Gap ~ t (Seconds)

FIGURE 1



Modification of charts by F. Thorndike, *Bell System Technical Journal*, (October 1926) and H. Working, *A Guide to Utilization of the Binomial and Poisson Distributions*, Stanford University Press, 1943.

FIGURE 2

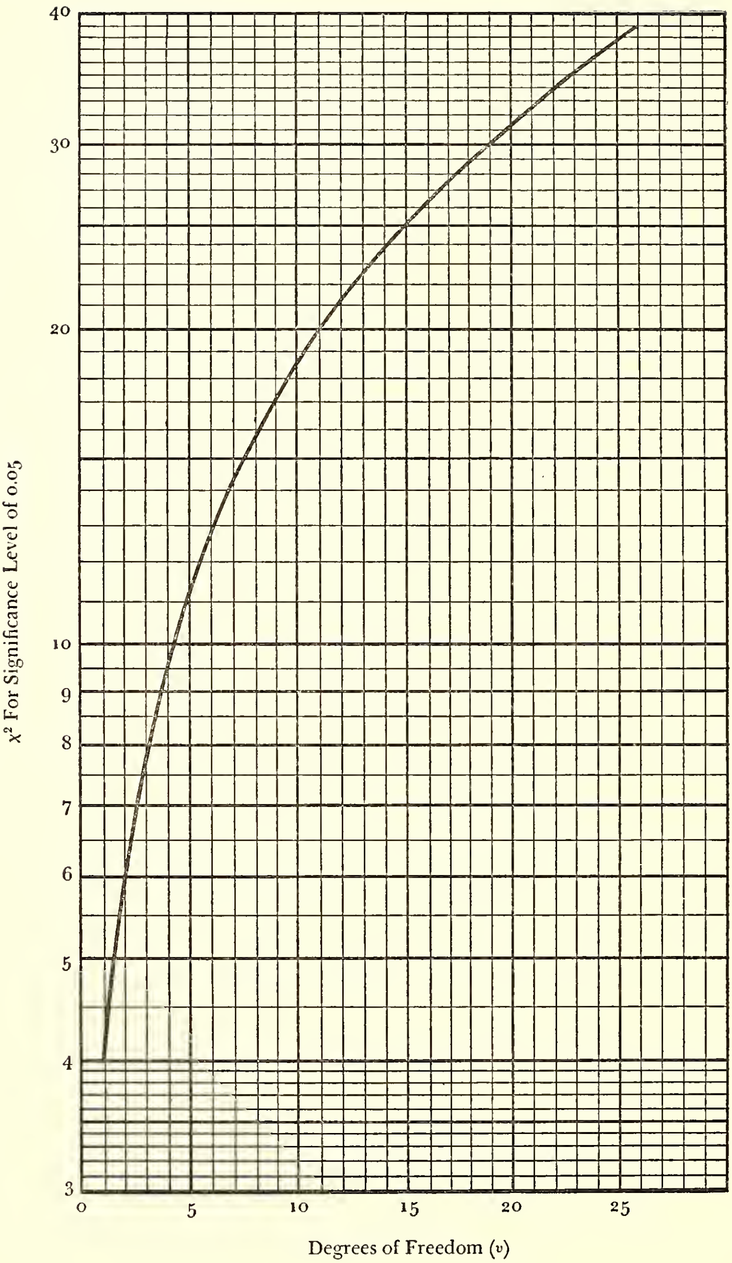

40
30
20
10
9
8
7
6
5
4
3
χ^2 For Significance Level of 0.05
0
5
10
15
20
25
Degrees of Freedom (v)

FIGURE 3

APPENDIX A

Greek Alphabet

Name	*Lower Case*	*Upper Case*
Alpha	α	Α
Beta	β	Β
Gamma	γ	Γ
Delta	δ	Δ
Epsilon	ϵ	Ε
Zeta	ζ	Ζ
Eta	η	Η
Theta	θ	Θ
Iota	ι	Ι
Kappa	κ	Κ
Lambda	λ	Λ
Mu	μ	Μ
Nu	ν	Ν
Xi	ξ	Ξ
Omicron	ο	Ο
Pi	π	Π
Rho	ρ	Ρ
Sigma	σ	Σ
Tau	τ	Τ
Upsilon	υ	Υ
Phi	ϕ	Φ
Chi	χ	Χ
Psi	ψ	Ψ
Omega	ω	Ω

APPENDIX B

Fundamental Concepts

Permutations and Combinations

The subject of permutations may be easily illustrated by the example of code words. Consider that 3 cards marked A, B, C are available and are to be used to form as many 3-letter code words as possible. The result will be:

A	B C	A B C
	C B	A C B
B	A C	B A C
	C A	B C A
C	A B	C A B
	B A	C B A

In each case there will be 3 choices for the first letter, 2 choices for the second letter, and 1 choice for the third letter. From this it follows that:

$$
\begin{aligned}
P_n^n &= \text{Permutations of } n \text{ things taken } n \text{ at a time} \\
&= n(n-1)\,(n-2) \cdot \cdot \cdot \cdot 3 \cdot 2 \cdot 1 \\
&= n! = \text{factorial } n \\
P_3^3 &= 3 \cdot 2 \cdot 1 = 6 \text{ which checks the empirical result}
\end{aligned}
$$

If there are 5 cards marked A, B, C, D, E and it is desired that 3-letter code words be formed, there will be 5 choices for the first letter, 4 choices for the second, and 3 for the third:

$$
\begin{aligned}
P_3^5 &= \text{Permutations of 5 things taken 3 at a time} \\
&= 5 \cdot 4 \cdot 3 = 60 \\
&= \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{5!}{2!} = \frac{5!}{(5-3)!}
\end{aligned}
$$

For the general case of the permutations of n things taken m at a time

$$P_m^n = \frac{n!}{(n-m)!}$$

In these code words the order is important, for ABC is different from ACB. Consider, however, 3-card hands formed from the 5 cards A, K, Q, J, 10. If order were important, there would be 60 permutations as follows:

AKQ	AQJ	AJ10	KQJ	AKJ	AK10	AQ10	KJ10	KQ10	QJ10
AQK	AJQ	A10J	KJQ	AJK	A10K	A10Q	K10J	K10Q	Q10J
KAQ	QAJ	JA10	QKJ	KAJ	KA10	QA10	JK10	QK10	JQ10
KQA	QJA	J10A	QJK	KJA	K10A	Q10A	J10K	Q10K	J10Q
QAK	JAQ	10AJ	JKQ	JAK	10AK	10AQ	10KJ	10KQ	10QJ
QKA	JQA	10JA	JQK	JKA	10KA	10QA	10JK	10QK	10JQ

Note that each vertical group is P_3^3 while the whole array is P_5^3. In general, however, the order of the cards is unimportant. It is the particular group of cards that is important (represented by the 10 groups above). When order is unimportant the grouping is known as a combination.

$$C_3^5 = \begin{matrix}\text{Combinations of 5 things}\\ \text{taken 3 at a time}\end{matrix} = \frac{\text{Permut. of whole array}}{\text{Permut. within combination}}$$

$$= \frac{P_3^5}{P_3^3} = \frac{60}{6} = 10$$

$$C_m^n = \frac{P_m^n}{P_m^m} = \frac{n!}{(n-m)!\,m!}$$

Factorial Zero

For clarity in certain problems it has been found convenient to define the factorial zero as follows:

$$(n-1)\,! = \frac{n!}{n}$$

let: $n = 1$

then: $(1-1)\,! = \frac{1\,!}{1}$

$$0\,! = 1$$

Laws of Probability

The following are two important laws concerning probability:

1. *Total Probability*
   If two events, A and B, are mutually exclusive (if A occurs, B cannot occur and vice versa) the total probability that one of these events will occur is:
   $$P(\text{A or B}) = P(\text{A} + \text{B}) = P(\text{A}) + P(\text{B})$$
2. *Joint Probability*
   If two events, A and B, are independent (the occurrence of one has no influence on the other) the probability that both will occur together is:
   $$P(\text{AB}) = P(\text{A and B}) = P(\text{A})P(\text{B})$$

APPENDIX C

The Binomial Distribution

Consider a population in which each item may possess one of two mutually exclusive characteristics (head or tail, good or bad, 0 or 1, etc.)

Let: p = probability of occurrence of characteristic A
$q = (1 - p)$ = probability of occurrence of characteristic B (non-occurrence of A)

Then, by the law of total probability,
$P(\text{A or B}) = p + q$

Suppose that a sample of n items is drawn from the population under the following conditions:

a. The size of the population is infinite. (This restriction insures that withdrawing the sample does not alter the relative proportion of A and B remaining in the population. The same result may be achieved with a finite population by drawing one item, replacing, stirring, drawing the next item, etc.)
b. The sample is selected from the population at random.

Then, by the law of joint probability:

the probability that all n items in the sample are A's $= p^n$

$$\left.\begin{array}{l}\text{the probability of } (n-1) \text{ A's and } 1 \text{ B} = p^{n-1}q \\ \text{the probability of } (n-2) \text{ A's and } 2 \text{ B's} = p^{n-2}q^2 \\ \quad\text{etc.} \\ \text{the probability of } 1\text{A and } (n-1) \text{ B's} = pq^{n-1} \\ \text{the probability of } n \text{ B's} = q^n \end{array}\right\} \left\{\begin{array}{l}\text{where the} \\ \text{order of} \\ \text{A's and} \\ \text{B's is sig-} \\ \text{nificant}\end{array}\right.$$

In general, for m A's and $(n-m)$ B's the probability $= p^m q^{n-m}$

where $m = 0, 1, 2, \ldots (n-1), n$

In drawing the sample, the order in which the A's occur can take on many possibilities, the number being equal to the combinations of n things taken m at a time $= C_m^n$

For instance, if the number of items in the sample is 5, there are $C_3^5 = 10$ ways in which a sample composed of 3A's and 2B's may be drawn. Each of these ways will have a probability of p^3q^2. Thus, by the law of total probability, the probability of 3 A's and 2 B's is:

$$10p^3q^2$$

or in general:

$$C_m^n p^m q^{n-m}$$

Considering the various possibilities of 0, 1, 2, ... n A's and n, $(n-1)$ . . . . 1, 0 B's, the total probability of occurrence of A's and B's is given by

$$P(A, B) = \sum_{m=0}^{n} C_m^n p^m q^{n-m}$$

The right hand member is equal to $(q + p)^n$ and hence this relationship is known as the binomial distribution.

$$\sum_{m=0}^{n} C_m^n p^m q^{n-m} = (p+q)^n$$

APPENDIX D

Derivation of the Poisson Distribution

Introduction: The Poisson Distribution is applicable to populations having the following properties:

a. The probability of occurrence of individuals having a particular characteristic is low.
b. The characteristic is a discrete variable.

The Poisson Distribution as a Limiting Case of the Binomial Distribution

Let n = number of items in sample

p = probability of occurrence of a particular characteristic E

$q = (1-p)$ = probability of non-occurrence of characteristic E

x = number of items in sample having characteristic E.

Then, from the binomial distribution:

$$P(x) = C_x^n p^x q^{n-x} = C_x^n p^x (1-p)^{n-x}$$

$$x = 0, 1, 2 \cdots n$$

Now let:

p be made indefinitely small

n be very large

$pn = m$, where m is finite and not necessarily small.

Then: $p = \dfrac{m}{n}$

$$P(x) = C^n_x \left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{n-x};\ x = 0, 1, 2, \cdots n$$

$$= \frac{n!}{x!\,(n-x)!} \left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{n-x}$$

$$= \frac{n!}{x!\,(n-x)!} \left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^n \left(1 - \frac{m}{n}\right)^{-x}$$

$$P(x) = \left[\frac{m^x}{x!}\right] \left[\left(1 - \frac{m}{n}\right)^n\right] \left[\frac{n!}{(n-x)!\,n^x\left(1 - \frac{m}{n}\right)^x}\right]$$

$$= [A]\,[B]\,[C]$$

Now, if $n \longrightarrow \infty$

$$\lim_{n\longrightarrow\infty} P(x) = \lim_{n\longrightarrow\infty} \left\{[A]\,[B]\,[C]\right\}$$

$$= \left[\lim_{n\longrightarrow\infty} A\right] \left[\lim_{n\longrightarrow\infty} B\right] \left[\lim_{n\longrightarrow\infty} C\right]$$

$$A = \frac{m^x}{x!}$$

$$\lim_{n\longrightarrow\infty} A = \frac{m^x}{x!}$$

$$B = \left(1 - \frac{m}{n}\right)^n$$

$$\lim_{n\longrightarrow\infty} B = e^{-m}$$

(See appendix E for proof)

$$C = \frac{n!}{(n-x)!\,n^x\left(1 - \frac{m}{n}\right)^x}$$

When n is very large, negligible error is introduced by representing $n!$ by one term of Stirling's formula. The same statement holds for $(n-x)!$

$$\therefore C = \frac{\sqrt{2\pi n}\ \ n^n\ \ e^{-n}}{\sqrt{2\pi\ (n-x)}\ (n-x)^{n-x}\ e^{-(n-x)}\left(1-\frac{m}{n}\right)^x n^x}$$

$$C = \frac{\sqrt{2\pi}n^{\frac{1}{2}}\ e^{-n}\ n^n}{\sqrt{2\pi}(n-x)^{\frac{1}{2}}\ (n-x)^{n-x}\ e^{-(n-x)}\left(1-\frac{m}{n}\right)^x n^x}$$

$$= e^{-x}\left(\frac{n-x}{n}\right)^{-\frac{1}{2}} \frac{n^{n-x}}{(n-x)^{n-x}} \frac{1}{\left(1-\frac{m}{n}\right)^x}$$

$$= e^{-x}\left(1-\frac{x}{n}\right)^{-\frac{1}{2}} \frac{1}{\left(\frac{n-x}{n}\right)^{n-x}} \frac{1}{\left(1-\frac{m}{n}\right)^x}$$

$$C = \left[e^{-x}\right]\left[1-\frac{x}{n}\right]^{-\frac{1}{2}} \frac{1}{\left(1-\frac{x}{n}\right)^{n-x}} \frac{1}{\left(1-\frac{m}{n}\right)^x}$$

$$= \left[e^{-x}\right]\left[1-\frac{x}{n}\right]^{x-\frac{1}{2}} \left[\frac{1}{\left(1-\frac{x}{n}\right)^n}\right]\left[\frac{1}{\left(1-\frac{m}{n}\right)^x}\right]$$

$$= \left[C_1\right]\left[C_2\right]\left[C_3\right]\left[C_4\right]$$

$$C_1 = e^{-x}$$

$$\lim_{n\to\infty} C_1 = e^{-x}$$

$$C_2 = \left(1-\frac{x}{n}\right)^{x-\frac{1}{2}}$$

$$\lim_{n \to \infty} C_2 = 1$$

$$C_3 = \frac{1}{\left(1 - \frac{x}{n}\right)^n}$$

$$\lim_{n \to \infty} C_3 = \frac{1}{e^{-x}}$$

(See Appendix E)

$$C_4 = \frac{1}{\left(1 - \frac{m}{n}\right)^x}$$

$$\lim_{n \to \infty} C_4 = 1$$

$$\begin{aligned}\lim_{n \to \infty} C &= \left[\lim_{n \to \infty} C_1\right]\left[\lim_{n \to \infty} C_2\right]\left[\lim_{n \to \infty} C_3\right]\left[\lim_{n \to \infty} C_4\right] \\ &= \left[e^{-x}\right]\left[1\right]\left[\frac{1}{e^{-x}}\right]\left[1\right] \\ &= 1\end{aligned}$$

$$\lim_{n \to \infty} P(x) = \frac{m^x}{x!} e^{-m}$$

Since the main body of this discussion assumes the existence of the conditions for the Poisson distribution the above equation may be written simply:

$$P(x) = \frac{m^x}{x!} e^{-m}$$

APPENDIX E

Derivation of Limit $\lim_{n\to\infty}\left(1 - \frac{m}{n}\right)^n$ (Used in the derivation of the Poisson distribution)

$$\text{Let} \quad n = \frac{1}{x}$$

$$\lim_{n\to\infty}\left(1 - \frac{m}{n}\right)^n = \lim_{\frac{1}{x}\to\infty}(1 - mx)^{\frac{1}{x}}; \text{ where } n = \frac{1}{x}$$

$$= \lim_{\frac{1}{x}\to\infty}\left[1 - \left(\frac{1}{x}\right)(mx) + \frac{\left(\frac{1}{x}\right)\left(\frac{1}{x} - 1\right)(mx)^2}{2!} - \frac{\left(\frac{1}{x}\right)\left(\frac{1}{x} - 1\right)\left(\frac{1}{x} - 2\right)(mx)^3}{3!} + \text{----}\right]$$

$$= \lim_{\frac{1}{x}\to\infty}\left[1 - m + \frac{m^2}{2!}(1 - x) - \frac{m^3}{3!}(1 - x)(1 - 2x) + \text{----}\right]$$

$$= 1 - m + \frac{m^2}{2!} - \frac{m^3}{3!} + \text{-----}$$

But expanding e^{-m} in a McLaurin series gives:

$$e^{-m} = 1 - m + \frac{m^2}{2!} - \frac{m^3}{3!} + \text{-----}$$

$$\therefore \lim_{n\to\infty}\left(1 - \frac{m}{n}\right)^n = e^{-m}$$

APPENDIX F

Mean, Variance, and Standard Deviation of Poisson Distribution

Mean

The mean of any continuous function is obtained by:

$$\mu = \frac{\int xf(x)dx}{\int f(x)dx}$$

(The mean may be considered as the distance to the center of gravity.)

For discrete functions the comparable function defining the mean is:

$$\mu = \frac{\Sigma xf(x)}{\Sigma f(x)}$$

When dealing with probabilities:

$$f(x) = P(x)$$

$$\int f(x)dx = 1$$

or

$$\sum_{x=0}^{\infty} P(x) = 1$$

$$\mu = \int xP(x)$$

or

$$\mu = \sum_{x=0}^{\infty} x\, P(x)$$

For the Poisson distribution

$$P(x) = \frac{m^x e^{-m}}{m!}$$

$$\mu = \sum_{x=0}^{\infty} \frac{x m^x e^{-m}}{m!}$$

$$= 0 + me^{-m} + \frac{2m^2 e^{-m}}{2!} + \frac{3m^3 e^{-m}}{3!} \cdots$$

$$= me^{-m}\left[1 + m + \frac{m^2}{2!} \cdots\right]$$

$$= me^{-m}e^{m}$$

$$= m$$

By definition, the variance σ^2 may be expressed:

$$\sigma^2 = \frac{\Sigma f(x)\ (x-\mu)^2}{\Sigma f(x)}$$

For the Poisson distribution

$$\sigma^2 = \frac{\Sigma(x-m)^2\ P(x)}{\Sigma P(x)} = \frac{\Sigma(x^2-2xm+m^2)\ P(x)}{1}$$

$$= \Sigma x^2 P(x) - 2m\Sigma xP(x) + m^2\Sigma P(x)$$

The last two terms reduce as follows:

$$-2m\Sigma xP(x) = -2m(m) = -2m^2$$

$$m^2\Sigma P(x) = m^2$$

The first term may be reduced by the following steps:

$$\Sigma x^2 P(x) = \Sigma\ [x(x-1) + x]\ P(x) = \Sigma x(x-1)\ P(x) + \Sigma xP(x)$$

$$= \Sigma x(x-1)\ P(x) + m$$

$$= \left[0 + 0 + \frac{2m^2\ e^{-m}}{2!} + \frac{6m^3\ e^{-m}}{3!} + \frac{12m^4\ e^{-m}}{4!} + \cdots\ \right] + m$$

$$= m^2 e^{-m}\left(1 + m + \frac{m^2}{2!} + \cdots\right) + m$$

$$= m^2 e^{-m}(e^m) + m = m^2 + m$$

$$\sigma^2 = m^2 + m - 2m^2 + m^2 = m$$

Standard Deviation

By definition the standard deviation, σ, is expressed by

$$\sigma = \sqrt{\sigma^2}$$

Thus, for the Poisson distribution

$$\sigma = \sqrt{m}$$

APPENDIX G

Notes on the χ^2 Test for Goodness of Fit

Let: f = Observed frequency for any group or interval

F = Computed or theoretical frequency for same group

Then, by definition:

$$\chi^2 \cong \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i} \tag{1}$$

where k = number of groups

Expanding:

$$\chi^2 = \sum_{i=1}^{k} \frac{f_i^2}{F_i} - \frac{2f_i F_i}{F_i} + \frac{F_i^2}{F_i}$$

$$= \sum_{i=1}^{k} \frac{f_i^2}{F_i} - 2 \sum_{i=1}^{k} f_i + \sum_{i=1}^{k} F_i$$

But by assumption in the fitting process:

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} F_i = n$$

where n = total number of observation

So that:

$$\chi^2 = \left(\sum_{i=1}^{k} \frac{f_i^2}{F_i} \right) - n \tag{2}$$

Either equation (1) or equation (2) may be used for purposes of computation. Usually (2) will simplify the amount of work involved.

The value of χ^2 obtained as above is then compared with the value from Figure 3 or from tables of χ^2. Such tables may be found in any collection of statistical tables, and relate the values of χ^2 and significance level with the degrees of freedom. The number of degrees of freedom, v, may be expressed:[20, 21]

$$v = G - 2$$

where G = number of groups

For this value of v to be valid, however, it is necessary that the theoretical number of occurrences in any group be at least 5. One writer[22] further stipulates that the total number of observations be at least 50. When the number of theoretical occurrences in any group is less than 5, the group interval should be increased. For the lowest and highest groups this may be accomplished by making these groups "all less than" and "all greater than" respectively.

20 Dixon, W. J., and Massey, F. J., Jr., *Introduction to Statistical Analysis*, McGraw-Hill, 1951, pp. 190-191.

21 Yule, G. Udny, and Kendall, M. G., *An Introduction to the Theory of Statistics*, 14th edition, Haffner Publishing Company, 1950, p. 475, p. 476.

22 Cramer, Harald, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 435.

APPENDIX H

Derivation of the Distribution of the Sum of Independent Poisson Distributions

Consider a population made up of two subpopulations A and B, each distributed according to the Poisson distribution.

For subpopulation A

$$P(x_a) = \frac{m^{x_a} e^{-m_a}}{x_a!}$$

For subpopulation B

$$P(x_b) = \frac{m^{x_b} e^{-m_b}}{x_b!}$$

If k items occur in a trial from the total population, there may be a mixture of x_a and x_b as follows:

$x_a = k;$	$x_b = 0$	$x_a + x_b = k$
$x_a = k - 1$	$x_b = 1$	$x_a + x_b = k$
$x_a = k - 2$	$x_b = 2$	$x_a + x_b = k$
$\cdots\cdots\cdots$		
$x_a = 2$	$x_b = k - 2$	$x_a + x_b = k$
$x_a = 1$	$x_b = k - 1$	$x_a + x_b = k$
$x_a = 0$	$x_b = k$	$x_a + x_b = k$

$$P(k) = P(x_a = k, x_b = 0) + P(x_a = k - 1, x_b = 1)$$

$$\cdots\cdots\cdots + P(x_a = 1, x_b = k - 1) + P(x_a = 0, x_b = k)$$

$$= \frac{m_a{}^k e^{-m_a} e^{-m_b}}{k!\ 0!} + \frac{m_a{}^{k-1} e^{-m_a} m_b e^{-m_b}}{(k-1)!\ 1!} + \frac{m_a{}^{k-2} e^{-m_a} m_b{}^2 e^{-m_b}}{(k-2)!\ 2!}$$

$$\cdots\cdots\cdots + \frac{m_a e^{-m_a} m_b^{k-1} e^{-m_b}}{1!\,(k-1)!} + \frac{e^{-m_a} m_b^k e^{-m}}{0!\,k!}$$

$$P(k) = e^{-m_a} e^{-m_b} \left\{ \frac{m_a^k}{k!} + \frac{k\, m_a^{k-1}\, m_b}{k(k-1)!} + \frac{k(k-1)\, m_a^{k-2}\, m_b^2}{k(k-1)\,(k-2)!\,2!} \right.$$

$$\left. + \cdots\cdots\cdots + \frac{k(k-1)\cdots 3\cdot 2\cdot 1\; m_a\, m_b^{k-1}}{k(k-1)\cdots 3\cdot 2\cdot 1\;(k-1)!} + \frac{m_b^k}{k!} \right\}$$

$$P(k) = \frac{e^{-(m_a + m_b)}}{k!} \left\{ m_a^k + k m_a^{k-1}\, m_b + \frac{k(k-1)\, m^{k-2}\, m_b^2}{2!} \right.$$

$$\left. + \cdots + k\, m_a\, m_b^{k-1} + m_b^k \right\}$$

$$P(k) = \frac{e^{-(m_a + m_b)}\,(m_a + m_b)^k}{k!}$$

When there are subpopulations A, B, ... , Z, by application of a similar argument the distribution for the whole population is found to be

$$P(k) = \frac{(m_a + m_b + \cdots + m_z)^k\, e^{-(m_a + m_b + \cdots + m_z)}}{k!}$$

The Probability Theory Applied To Distribution of Vehicles On Two-lane Highways

André Schuhl

Mr. Schuhl is engineer of bridges and roads in the French Ministry of Public Works. In France these engineers are graduated from the Polytechnic School and then from the National School for Bridges and Roads. Mr. Schuhl is also a graduate of the National Superior School of Electricity. During the war he studied hydroelectrical plants in Pyrénées and Massif Central. He joined the Free French Forces through Spain early in 1944 and served with the first Tactical Air Force of the Seventh Army Group as a French liaison officer to the American Army. He came to the United States in 1954 with Mission 149 of the O.E.C.E. to study traffic engineering and traffic control.

A theoretical study of conditions affecting the traffic of vehicles on a highway requires the constant use of probability theory. On the one hand, the distribution of vehicles in each lane is in part a matter of chance. On the other hand, whenever one studies the behavior of any large number of individuals, inevitable departures from the laws which apply to their totality are found by statistical analysis to follow certain empirical and relatively stable patterns. These are, in effect, special laws of probability.

Reprinted with permission from *Travaux,* January 1955, revised and updated by the author. Translated from the French.

In particular, Poisson's law, applying to rare events, has thus far been the chief theoretical instrument for dealing with problems of vehicular traffic on two or three lane highways, especially the problem of determining the distribution of such traffic both in time and in space. This law, as is well known, assigns the value $e^{-N\theta}\frac{(N\theta)^n}{n!}$ to the probability that there are "n" vehicles in a time interval θ chosen at random, N being the average number of vehicles in unit time. The probability that there are no vehicles in this interval is

$$e^{-N\theta}$$

While the results yielded by this formula agree well enough with actual observation when the traffic density is low (a few dozen cars per hour per lane), they differ widely from reality when the density is significantly larger. The reasons for this discrepancy between theoretical prediction and the observed data are clear enough. But without going into any analysis of physical causes, it seems obvious that the gap between theory and actuality can be considerably narrowed. It is the object of this paper to show how this can be done.

We shall suppose that the entire set of spacings between successive vehicles consists of a number of distinct parts or sub-sets, each having distinct mean values and each obeying some Poisson-type law. For simplicity we consider just two sub-sets. Let the number of spacings per unit time for these two sets be respectively

$$\gamma N \text{ and } (1-\gamma)N$$

with mean spacing-values of t_1 and t_2 seconds respectively, and with $t_1 < t_2$. We also suppose that the entire set of spacings is a set of random and independent elements.

Then the probability that there is no vehicle in an interval θ would be $e^{-\frac{\theta}{t_1}}$ if all spacings were in the first sub-set and $e^{-\frac{\theta}{t_2}}$ if they were all in the second sub-set.

Now the respective times covered by the two sub-sets are

$$\gamma N t_1 \text{ and } 1-\gamma N t_1,$$

and clearly

$$1-\gamma N t_1 = (1-\gamma) N t_2$$

Accordingly, the probability that there are no vehicles in an interval θ will be given by

$$P(\theta) = N\gamma t_1 e^{-\frac{\theta}{t_1}} + (1-\gamma) N t_2 e^{-\frac{\theta}{t_2}} \tag{1}$$

with the relation

$$N\gamma t_1 + N(1-\gamma) t_2 = 1,$$

which merely states that the N spacings cover unit time. Before proceeding further, it must be observed that the first set of spacings might apply to retarded vehicles which are prevented from passing by opposing traffic, and the second set to free-moving vehicles which are able to pass at will. As vehicles cannot be considered as mere points, two successive vehicles in the first set must necessarily be separated by a time interval having a positive lower bound ϵ. On the contrary, free-flowing vehicles having opportunities to pass may exhibit spacings equal to zero.

Hence the law of spacings which we use in practice is not formula (1), but rather the formula

$$P(\theta) = N\gamma(t_1-\epsilon) e^{-\frac{\theta-\epsilon}{t_1-\epsilon}} + N(1-\gamma) t_2 e^{-\frac{\theta}{t_2}} \tag{2}$$

We must now examine the practical usefulness of the laws embodied in formulas (1) and (2) in providing answers to various questions arising out of the study of traffic problems. They are considered in some detail in what follows.

First Question: Probability of a spacing longer than x.

Suppose that the vehicles in our set are represented by points on a straight line and that $p(x)dx$ is the probability of a spac-

ing between x and $x + dx$. What then is the probability $P(\theta)$, that an interval θ chosen at random on the straight line, contains no points, all choices of θ on the line being equally probable?

Now the probability that the initial point A of the interval $AB = \theta$ will be in a spacing lying between x and $x + dx$ is

$$kxp(x)dx,$$

k being defined by the condition

$$\int_0^\infty kxp(x)dx = 1,$$

which simply states that the point A necessarily falls in a spacing which lies between O and ∞. The value of k is given by

$$k = \frac{1}{\int_0^\infty xp(x)dx} = \frac{1}{x_m} = N,$$

where x_m is the mean value of x.

If the interval $AB = \theta$ contains no points, L being the first point preceding A, and M the first one following A, with $LM = x$, it follows that $x \geqslant \theta$ and that the initial point A lies between L and P, where $LP = x - \theta \geqslant O$. Now the probability that the interval $AB = \theta$ contains no points is the same as the probability that its initial point A lies between L and P, and this probability is found from the above to be

$$k\left(\frac{x-\theta}{x}\right) xp(x)dx = N(x-\theta)p(x)dx$$

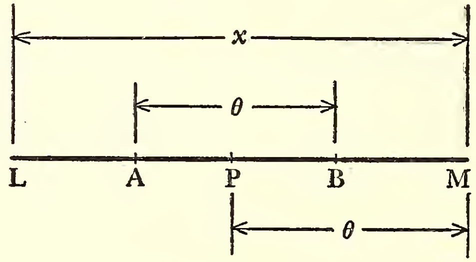


FIGURE 1

Hence upon integration over all values of x between θ and ∞, we have

$$P(\theta) = \int_{\theta}^{\infty} N(x-\theta)p(x)dx = \int_{\theta}^{\infty} N\,dx \int_{x}^{\infty} p(x)dx \tag{3}$$

From this we find by two successive differentiations, that

$$p(x) = \frac{1}{N}\frac{d^2 P(x)}{dx^2}$$

Using this result, we find by taking the value of $P(x)$ from (1) that in this case

$$p(x) = \frac{\gamma}{t_1} e^{-\frac{x}{t_1}} + \frac{1-\gamma}{t_2} e^{-\frac{x}{t_2}} \tag{4}$$

while the total probability of a spacing longer than x is given by

$$\int_{x}^{\infty} p(x)dx = \gamma\, e^{-\frac{x}{t_1}} + (1-\gamma)e^{-\frac{x}{t_2}} \tag{5}$$

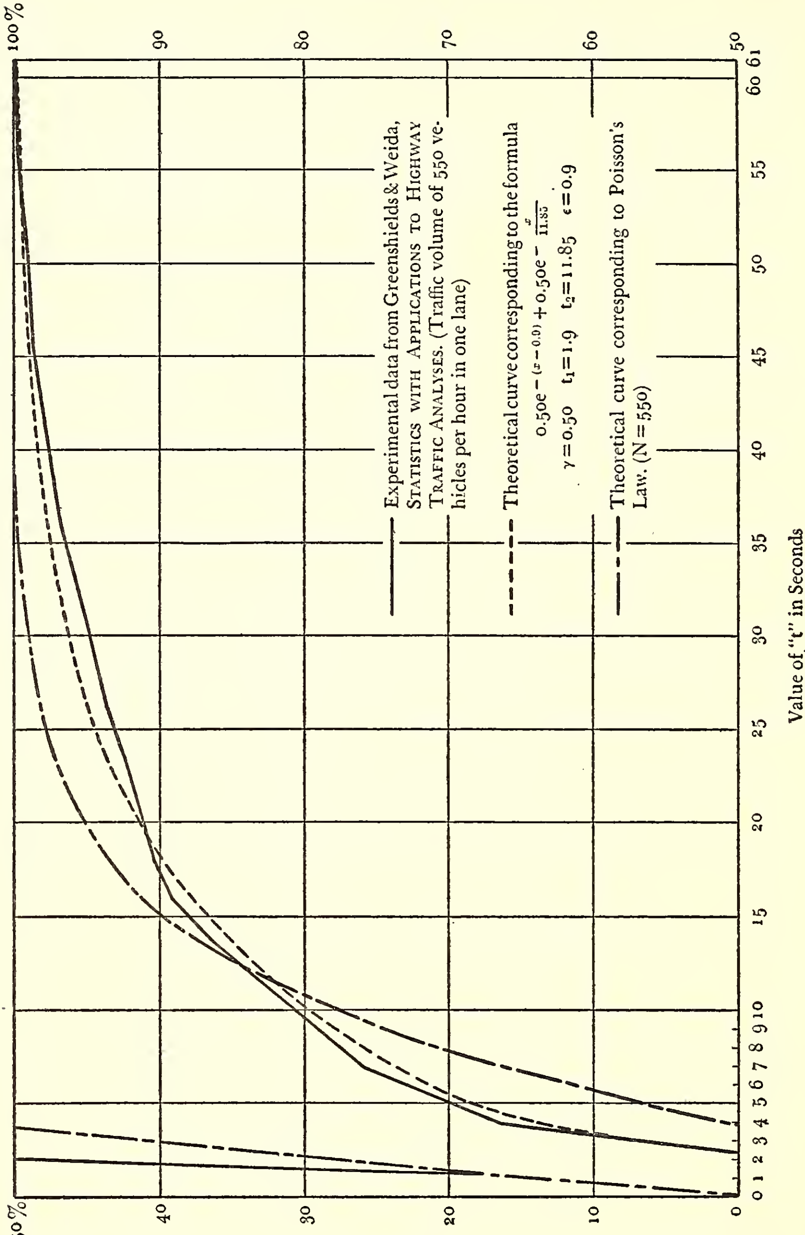
If on the other hand we take the value of $P(x)$ furnished by (2), then

$$p(x) = \frac{\gamma}{t_1-\epsilon} e^{-\frac{x-\epsilon}{t_1-\epsilon}} + \frac{1-\gamma}{t_2} e^{-\frac{x}{t_2}} \tag{6}$$

and

$$y = \int_{x}^{\infty} p(x)dx = \gamma\, e^{-\frac{x-\epsilon}{t_1-\epsilon}} + (1-\gamma)e^{-\frac{x}{t_2}} \tag{7}$$

It will be seen by examining the graphs of Figure 2 which has been taken from *Statistics with Applications to Highway Traffic Analyses* by Greenshields and Weida, that there is good agreement between the observed data and those furnished by formula (7). The legend on the diagram gives the actual numerical values of the various parameters used in this particular case. (See Appendix, p. 73.)



100%
90
80
70
60
50
50%
40
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10
15
20
25
30
35
40
45
50
55
60 61
Value of "t" in Seconds
Experimental data from Greenshields & Weida, STATISTICS WITH APPLICATIONS TO HIGHWAY TRAFFIC ANALYSES. (Traffic volume of 550 vehicles per hour in one lane)
Theoretical curve corresponding to the formula
$0.50e^{-(t-0.9)} + 0.50e^{-\frac{t}{11.85}}$
$\gamma=0.50$ $t_1=1.9$ $t_2=11.85$ $\epsilon=0.9$
Theoretical curve corresponding to Poisson's Law. (N=550)

FIGURE 2

Figure 3 gives another example, studied on a French road, showing how actual data are approximated by the theoretical formula. There were 312 spacings observed in a time period of 52 minutes, 45 seconds. Opposing traffic amounted to only 12 percent of the total flow.

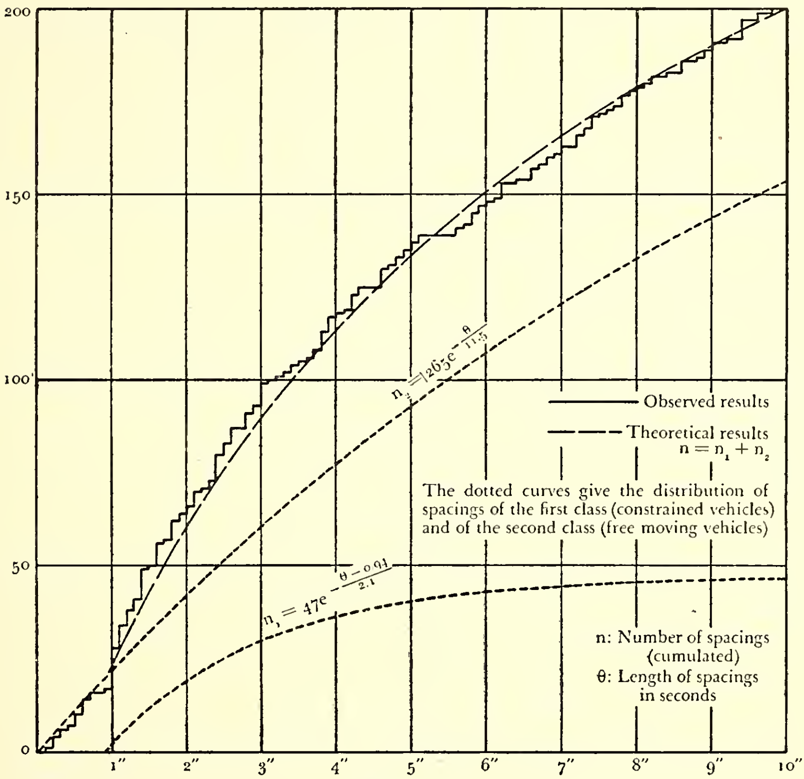


FIGURE 3

Second Question: Probability that an interval θ, taken at random, contains no vehicles, but is bounded by a vehicle on one side.

First, we clarify the significance of the iterated integrals occurring in formula (3).

If $AB = \theta$ is placed at random on the spacing $LM = x$, AB can be empty only if A and B are between L and M.

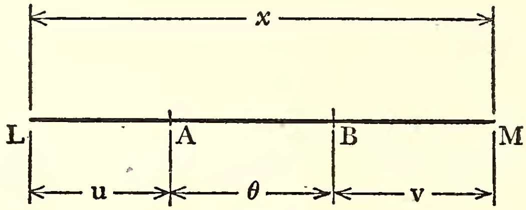


FIGURE 4

Accordingly, $x \geqslant \theta$, and Figure 4 shows that the various cases in which AB contains no points, may be obtained by letting L range over all positions preceding A, and M over all positions following B.

If B lies between M and M', where $MM' = dv$, and if kdv is the probability of this happening, then the probability that LM lies between u and $u + du$ is the probability of a spacing

$$u + \theta + v < x \leqslant u + du + \theta + v,$$

that is,

$$p(u + \theta + v)du$$

Therefore the total probability that L precedes A, M being given on dv, is

$$k\,dv \int_0^\infty p(u + \theta + v)du$$

Also the total probability that L precedes $A(u \geqslant o)$ and that M follows $B(v \geqslant o)$ is

$$\int_0^\infty k\,dv \int_0^\infty p(u + \theta + v)du$$

If we make the substitution

$$u + \theta + v = x,$$

this can be written in the form

$$\int_{o}^{\infty} k\,dv \int_{\theta+v}^{\infty} p(x)dx$$

By the further substitution

$$\theta + v = y,$$

we find

$$\int_{\theta}^{\infty} k\,dy \int_{y}^{\infty} p(x)dx = k \int_{\theta}^{\infty} p(x)dx \int_{\theta}^{\infty} dy,$$

or

$$k \int_{\theta}^{\infty} (x-\theta)p(x)dx$$

where k is determined by the condition that $P(O) = 1$, that is, $k = N$, and we recover formula (3).

The upshot of this argument is the demonstration that there are several distinct definitions of the probability that a given time interval is empty. In particular our second question can now be answered, for the probability that a time interval between θ and $\theta + d\theta$ is empty *but bounded on one side by a vehicle*, is found from the above results by taking $MB = v = O$, while L of course precedes A. In fact, if $v = O$ we obtain from formula (3) by differentiation

$$-\frac{dP}{d\theta}\,d\theta = Nd\theta \int_{\theta}^{\infty} p(x)dx$$

We shall hereafter write

$$-\frac{dP}{d\theta} = J(\theta)$$

Third Question: Probability that an interval θ, taken at random, contains exactly n vehicles.

If x_1, x_2, -----x_n are the distances of the various points

(vehicles) interior to θ measured from the initial point A of the interval, we can use the preceding results to find the probabilities that the intervals $\overline{Ax_1}$, $\overline{x_1x_2}$, ----- $\overline{x_{n-1}\, x_n}$ are empty but bounded on the right by a point (vehicle), namely x_1 for the first interval, x_2 for the second, and x_n for the last.

Consider the interval $\overline{Ax_1}$. The probability that it has the required property is

$$\int_o^\theta J(x_1)dx_1$$

Now the point x_1 must be followed by a spacing smaller than $\theta - x_1$, that is, the corresponding interval must be empty but bounded on the right by a point, viz. x_2. The probability of this happening is

$$\int_o^\theta J(x_1)dx_1 \int_{x_1}^\theta J(x_2-x_1)dx_2$$

We continue in this way, up to the last interval $\theta - x_n$, which is empty but not bounded on the right. Hence the desired probability is

$$\int_o^\theta J(x_1)dx_1 \int_{x_1}^\theta J(x_2-x_1)dx_2 \text{ ----- } \int_{x_{n-1}}^\theta J(x_n-x_{n-1})dx_n\, P(\theta-x_n)$$

It is this formula which gives the value $e^{-N\theta}\dfrac{(N\theta)^n}{n!}$ when

$P(\theta) = e^{-N\theta}$.

Fourth Question: Probable delay in waiting for an empty interval θ.

This question comes up whenever a vehicle in one lane wishes to pass another car ahead of it, also whenever a vehicle enters upon or cuts across a main highway. We first determine the probability that an interval of length l, whose initial point is chosen at random, contains no void of length θ.

This is obviously a function $F(l, \theta)$ which steadily decreases

as l increases, θ being given. The decrease in $F(l, \theta)$ due to an increase dl is the probability that the interval $l + dl$ contains a void θ, while l itself contains no void. It is, in other words, the probability that the interval l contains a void $\theta - dl$ at its extremity, dl itself being empty. Consequently it is the product of the probabilities that $l + dl - \theta$ contains no void of length θ and that $\theta - dl$ is empty but bounded on one side by a vehicle, provided we neglect all infinitesimals of order higher than that of $F(l - \theta, \theta)J(\theta)dl$. We can therefore write

$$\frac{\delta F(l, \theta)}{\delta l} = - F(l-\theta, \theta)J\theta$$

This relation enables us to define $F(l, \theta)$ step by step once we know it on an interval of minimum length θ.

Now

$$F(l, \theta) = 1 \qquad \text{for} \qquad O \leqslant l < \theta$$

and from this we obtain

$$\frac{\delta F}{\delta l} = - J(\theta) \text{ and } F = K - J(\theta)l \text{ for } O \leqslant l < 2\theta$$

Since obviously

$$F(\theta, \theta) = 1 - P(\theta),$$

we see that

$$K = 1 - P(\theta) + J(\theta)\cdot\theta,$$

and

$$F(l,\theta) = 1 - P(\theta) - J(\theta)(l-\theta)$$

Continuing in this manner to

$$n\theta \leqslant l < (n+1)\theta$$

we find that

$$F(l, \theta) = 1 + \sum_{1}^{n}(-1)^n \left[\frac{P.J^{n-1}}{(n-1)!}(l-n\theta)^{n-1} + \frac{J^n}{n!}(l-n\theta)^n\right].$$

We are now in a position to answer the fourth question.

The probability that the first empty interval of minimum length θ begins at a distance l from the origin, is equivalent to the probability that the interval l contains no void, while the interval θ which immediately follows it is empty but bounded on one side, and this is

$$F(l, \theta)J(\theta)dl,$$

or, from our definition

$$-\frac{\delta F(l+\theta, \theta)}{\delta l}\,dl.$$

From this it follows that

$$\int_0^\infty F(l, \theta)J(\theta)dl = F(\theta, \theta) = 1 - P(\theta)$$

which means, as was to be expected, that the total probability that a void θ begins at a positive distance from the origin is the complement of $P(\theta)$, which is the probability that the void θ begins precisely at the origin.

The probable delay in waiting for a void of minimum length θ can therefore be written as

$$\rho = \int_0^\infty lF(l, \theta)J(\theta)dl = -\int_0^\infty l\,\frac{\delta F(l+\theta, \theta)}{\delta l}\,dl$$

$$= -\Big[F(l+\theta, \theta)l\Big]_0^\infty + \int_0^\infty F(l+\theta, \theta)dl$$

The first term vanishes at both limits, and we also know from the properties of the function F, that

$$F(l+\theta, \theta) = -\frac{1}{J(\theta)}\,\frac{\delta F(l+2\theta, \theta)}{\delta l}$$

Therefore we can write the second term in the form

$$\rho = \frac{F(2\theta, \theta)}{J(\theta)} = \frac{1-P(\theta)}{J(\theta)} - \theta$$

and this is the answer to our question.

It may be noted that this probable value ρ takes account of the cases where $l = O$, whose probability is $P(\theta)$.

We must bear in mind that the void θ required for a passing maneuver depends upon whether or not the vehicle in question has to slow down before passing. Taking for example θ_1 to be the void necessary for an unretarded passing, and $\theta_2 > \theta_1$ to be the void required for a retarded passing, the probable waiting delay is given by

$$\rho = \frac{1 - P(\theta_1)}{\mathcal{J}(\theta_2)} - \theta_2 \frac{1 - P(\theta_1)}{1 - P(\theta_2)}$$

Fifth Question: Mean number of vehicles contained in a group. Probability of finding a group with a given number of vehicles.

This question is of interest in the study of the distribution of spacings. We shall call a collection of vehicles a group if successive vehicles contained in it are separated by spacings shorter than a given value t, while successive groups are separated by spacings longer than t. Now on the basis of the preceding developments, we know that the probability of a spacing longer than t is $\frac{\mathcal{J}(t)}{N}$, and that the probable number of spacings per unit time is therefore $\mathcal{J}(t)$ where N is the entire number of spacings.

This is also, as we see, the probable number of groups of vehicles per unit time, and the probable mean number of vehicles in a group will be

$$\mu = \frac{N}{\mathcal{J}(t)}$$

Moreover, the probability that there are exactly n vehicles in a group is

$$\frac{\mathcal{J}(t)}{N}\left[1 - \frac{\mathcal{J}(t)}{N}\right]^{n-1}$$

a result which remains valid for $n = 1$.

These considerations afford a means of determining the law of distribution of spacings (by studying the composition of the various groups) of vehicles which pass a check point.

Conclusion

It appears that the results presented in this paper can lead to a considerably closer agreement between theory and observation than has been obtained heretofore. Their interest consists in the fact that in spite of the somewhat complicated form of the law we have chosen for the probability of spacings between successive vehicles, the values which it gives in answer to our five fundamental questions are comparatively simple. They involve only the functions P, J, and p. Of course one could, if one preferred, start with a purely empirical law, as for example the empirical curve of Figure 2 which is simply $\frac{J(\theta)}{N}$, and from this obtain the curves for P and p by graphical integration or differentiation. Our questions could thus be answered without any assumed law of probability.

It is hoped that these results will be useful to those interested in traffic research, who recognize the growing importance of statistical studies in this field. They may also throw light on certain extensions of Poisson's Law.

APPENDIX

For readers interested in the goodness of fit of the time-spacing distribution represented by the formula here developed, the following chi-square (χ^2) test data are submitted.

The author has found it desirable to replace the value of the various parameters given in the text by the following values:

$t_1 = 1.98$ $\quad t_2 = 13.16$ $\quad \gamma = 0.583$

$\epsilon = 0.81$ $\quad 1 - \gamma = 0.417$

These lead to the formula

$$0.583e^{-\frac{(\theta-0.81)}{1.17}} + 0.417e^{-\frac{\theta}{13.16}}$$

in percent of the total number of spacings (Fig. 1) and

$$385 \times 10^{-0.37\,(\theta-0.81)} + 275 \times 10^{-0.033\theta}$$

in number of spacings (table).

Col. 1 gives the limits of the intervals in which the spacings are classed.

Col. 2 the theoretical number of spacings n, corresponding to the entries in Column 1.

Col. 3 the theoretical number of spacings f_t comprised between times given by Column 1.

Col. 4 the observed number f_o corresponding to f_t of Column 3.

Col. 5 the deviation $f_t - f_o$.

Col. 6 the values of $\frac{(f_t-f_o)^2}{f_t}$. The sum of this column, namely

$$\sum \frac{(f_t-f_o)^2}{f_t}, \text{ is } \chi^2.$$

FITTING OF THE LAW

$$n = 385 \times 10^{-0.37(\theta-0.81)} + 275 \times 10^{-0.033\theta}$$

(Chi-Square Test)

θ	n	f_t	f_o	$f_t - f_o$	$\frac{(f_t-f_o)^2}{f_t}$
0	660				
		77.66	78	− 0.34	0.00
1	582.34				
		206.42	207	− 0.58	0.00
2	375.92				
		97.39	94	+ 3.39	0.12
3	278.53				
		50.19	58	− 7.81	1.22
4	228.34				
		29.43	24	+ 5.43	1.00
5	198.91				
		19.97	17	+ 2.97	0.44
6	178.94				
		15.41	23	− 7.59	3.74
7	163.53				
		12.95	11	+ 1.95	0.29
8	150.58				
		21.80	18	+ 3.80	0.66
10	128.78				
		18.26	23	− 4.74	1.23
12	110.52				
		15.61	20	− 4.39	1.23
14	94.91				
		13.38	16	− 2.62	0.51
16	81.53				
		11.49	7	+ 4.49	1.75
18	70.04				
		9.88	6	+ 3.88	1.52
20	60.16				
		8.48	4	+ 4.48	2.37
22	51.68				
		7.28	6	+ 1.28	0.22
24	44.40				
		6.27	6	+ 0.27	0.01
26	38.13				
		12.05	10	+ 2.05	0.35

FITTING OF THE LAW (*Continued*)

θ	n	f_t	f_o	$f_t - f_o$	$\frac{(f_t - f_o)^2}{f_t}$
31	26.08				
		8.24	11	— 2.76	0.92
36	17.84				
		5.64	8	— 2.36	0.99
41	12.20				
		3.86	6	— 2.14	1.19
46	8.34				
		2.63	1	+ 1.63	1.01
51	5.71				
		3.04	4	— 0.96	0.30
61	2.67				
		1.42	1	+ 0.42	0.12
71	1.25				
		1.25	1	+ 0.25	0.05
00	0.00				

chi-square = 21.24

Degrees of Freedom = 25—5=20